MATTER, MOTION AND ELECTRICITY

A Modern Approach to General Physics

BY
HENRY DE WOLF SMYTH
Chairman, Department of Physics
Princeton University

AND

CHARLES WILBUR UFFORD
Professor of Physics, Allegheny College

FIRST EDITION

McGRAW-HILL BOOK COMPANY, INC.
NEW YORK AND LONDON
1939

PRINTED IN THE UNITED STATES OF AMERICA

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE

Several years ago the Department of Physics at Princeton decided to reorganize its introductory program. It was felt that some recognition should be given to the improved courses in physics in school and the increasing number of students taking such courses. But it was realized that it is not yet possible to feel sure that a school course has laid adequate foundations for advanced work. The problem was to plan a course that reviewed fundamental principles sufficiently to ensure a solid base on which to build and yet avoided too exact a repetition of the conventional beginners' course. Evidently a novel approach was needed. A little thought suggested that such an approach was already at hand, in fact pressing for adoption, the approach offered by the developments of atomic physics in the last forty years. Such an approach gives a unity to the subject; it makes it possible, even necessary, to introduce modern developments early and often; and, finally, it avoids a repetition of the school course.

We never intended to write a textbook. But once convinced of the wisdom of our fundamental plan, we had no choice because we could find no book that followed such a plan. The book has grown gradually over a period of years, and during this period another guiding principle has evolved. This principle is that it is better to treat a few topics in an adult analytical way as completely as the preparation of the student allows rather than to cover the whole field. We are more anxious that a good student should understand the kind of reasoning and judgment used in physics than that he should read a description of all the results obtained, from rotor ships to cosmic rays. Actually, we have had to compromise this principle with the needs of our students, and we fear that many topics are too cursorily treated. Nevertheless, we have chosen three major experiments for unusually complete treatment, that of Cavendish on gravity, that of Millikan on the electronic charge, and that of J. J. Thomson on positive rays. The last of these was chosen partly for local

reasons, but also because it gives so direct a comparison between the results of mathematical reasoning and the results of experiment. We have omitted hydrostatics and geometrical optics entirely because experience shows that students learn these topics in school about as well as they can learn them in a first-year college course. Similar considerations might have ruled out some of the material on specific heats, but it was included as preparation for some of the kinetic-theory results. The thermodynamics usually included in an elementary course we have omitted for the opposite reason. We think the concepts are too difficult to be grasped in the short time that can be allotted to them. Other topics, such as most of sound, we have omitted as being somewhat apart from the main current of the subject, particularly from the atomic-physics point of view we have adopted.

After considerable hesitation, we have decided to use the meter-kilogram-second system of units throughout and, of course, in keeping with that choice, the practical units in electricity. This system is discussed in a separate introductory note. We have not used it heretofore in the course from which this book arose, but we believe its advantages justify its use without preliminary trial.

This book is not intended for self-instruction. It needs to be supplemented with lecture demonstrations, explanations of problems, and descriptions of practical applications. It is primarily concerned with the principles of physics and the method by which they are discovered.

The course from which this book arose was a group enterprise, and the book was first planned as a similar project. In the end, practically all the text was written by one of us (H.D.S.), except part of Chap. IV, which was written by E. U. Condon, and Chap. XXVII, which was largely written by L. N. Ridenour. But others at Princeton, particularly L. A. Turner and G. P. Harnwell, have also contributed many ideas and suggestions. We are extremely grateful to them and to several colleagues in other institutions who have furnished the photographs from which some of the cuts were made.

Princeton, N. J.,
Allegheny, Pa.,
August, 1939.

H. D. Smyth,
C. W. Ufford.

CONTENTS

INTRODUCTORY NOTE ON UNITS

The dual nature of physics is sometimes fascinating and sometimes exasperating. We may rejoice when highly abstract logical reasoning shows a result that is confirmed by practical measurement but our pleasure turns to grief when some definition, exquisite in its logic, proves almost useless in describing experimental observations. It is this duality, this constant compromise between logic and experience, that makes it so difficult to set up a completely satisfactory system of units appropriate to all parts of the subject. As a result, many different systems of units have grown up and most books on elementary physics use at least three or four of them. We are going to emphasize one that is not commonly used because we feel that its advantages are more nearly universal than those of any other system with which we are familiar.

Familiar Units

In the experiences of everyday life, many different physical quantities are encountered and expressed numerically. The inch, foot, and mile and to a less extent the centimeter, meter, and kilometer are familiar units in which lengths are expressed. We all measure time in seconds, minutes, hours, and so on. The pound and perhaps the gram and kilogram are familiar as units of quantity of material—what the physicist defines more precisely and calls mass. Power units are used in automobile advertising and in the electrical appliance industry though the innocent consumer may not recognize that a five-hundred-watt light bulb is described equally well as a 0.67-horsepower light bulb or that a hundred-horsepower car is also a seventy-five-kilowatt one. Other electrical units, the volt and ampere, are universally used, even if not understood.

Relations between Units

In most of our ordinary experiences, we are not conscious of any relation between these units. We think of each as expressing

a different kind of quantity and only think of relations between these different quantities in a vague sort of way. We know that an automobile of a hundred horsepower will burn more gallons of gasoline and travel at a speed of more miles per hour than one of ten horsepower, but we do not try to say exactly how many gallons per hour equal how many horsepower equal how many miles per hour. Perhaps the only units that we use more precisely are the electrical ones of power and energy. We do know that if we burn ten hundred-watt bulbs for an hour we have used up one kilowatt-hour of electrical energy and should pay the power company accordingly. But it is just this kind of precise relation that a physicist likes, and he wishes to use units that are logically related in all branches of the subject.

The Metric System

The first great step in the development of a logical system of related units was the introduction of the metric system at the beginning of the nineteenth century. In this system, the meter was set up as the unit of length. It was supposed to be one ten-millionth of the distance from the equator to the pole along the meridian through Paris. Other units of length, the centimeter, millimeter, and kilometer, were, respectively, $\frac{1}{100}$, $\frac{1}{1000}$, and 1,000 meters. Area was then measured in square meters, square centimeters, etc., and volumes were measured in cubic meters, cubic centimeters, etc. One cubic centimeter of water was then taken as a standard of mass, and its mass was called the gram. Because of the practical difficulties of applying these exact definitions exactly, they were superseded by definitions in terms of the length of a standard meter of platinum-iridium alloy and the mass of a standard kilogram of the same material. Both of these standards are kept at the International Bureau of Weights and Measures near Paris. The standard of time was chosen as the mean solar second.

The Centimeter-gram-second System

When these units began to be generally used by physicists, it seemed that the size of the centimeter and gram were on the whole more convenient than the meter and kilogram. A whole system of units was built up on this basis. The unit of velocity was the centimeter per second; the unit of acceleration was the

centimeter per second per second. Then a unit of force was defined as a force that would give to a mass of one gram an acceleration of one centimeter per second per second. This force was called one dyne. Then the work done by a force of one dyne acting through a distance of one centimeter was defined as the unit of work and called one erg, and so on. The system to which these units belong is called the centimeter-gram-second or c.g.s. system. It has many advantages but has the disadvantage that the units of force and energy are extremely small. For example, the work done in raising a pencil one inch is about thirty thousand ergs.

Electrical Units

While the c.g.s. system was being developed in scientific work and units of electric charge and current defined in it, practical telegraphers were finding it necessary to express electric currents and electromotive forces in some sort of units. A certain type of battery known as the Daniell cell was almost universally used as a source of electromotive force and was therefore chosen as a unit. Fortunately before this was standardized, it was realized that it was equal almost exactly to one hundred million, or 10^8, times the unit reached by one method of logical deduction (the electromagnetic) from the c.g.s. system of mechanical units. Consequently the volt or practical unit of electromotive force was defined in such a way as to make this ratio hold exactly. At the same time, corresponding units of current, charge, and resistance were defined. These units and others derived from them constitute the so-called practical system of electric units which is in universal use today. In this practical system of electrical units, the unit of energy is the watt-second or joule which is just ten million (10^7) ergs or c.g.s. units.

Here is a good illustration of the difficulty mentioned in the first paragraph. If we were dealing with batteries and currents in a purely empirical way and not worrying about the relations between electromotive force and current or the mechanical forces set up between currents, we could always use the practical system of units. They need not even be related to each other any more simply than a cubic foot is to a gallon (= 0.134 cu. ft.) in the "practical" English system of units of length and volume. On the other hand, if we were concerned only with the logical

MATTER, MOTION, AND ELECTRICITY

CHAPTER I

ATOMS AND MOLECULES

1. The natural sciences attempt to study objectively the events in the world of nature and the relations between these events. As a matter of convenience, it has become customary to classify these natural phenomena into several groups and give names to the studies of particular groups, such as physics, chemistry, biology, and so on. In many cases, this classification of phenomena is quite arbitrary. The actual problems with which physics is concerned today are often of vital interest to chemists, and a full understanding of modern physics is impossible without some familiarity with elementary chemistry. The physics of today is much concerned with atoms and molecules and the structure of matter, problems once considered the peculiar province of chemistry. Similarly a modern general chemistry text has sections about electrical discharge in gases, x-rays, and other fields of physics. This overlapping of the two sciences may embarrass the educator eager to divide all knowledge into packages or "credits" of so many hours per week per term, but it is of great advantage to the two sciences and to the whole body of knowledge of which they form a part. In this common ground of physics and chemistry, we find a natural starting point for the treatment of either subject; it is by no means a traditional starting point for the development of the subject of physics, but it is one that has many advantages. Because of them, we shall begin this book with a study of atoms and molecules. By doing so, we shall be able to interpret and correlate many physical phenomena which otherwise would have to be studied as isolated and unexplained wonders of nature.

2. The fundamental question of the structure of matter was clearly formulated by the philosophers of ancient Greece. Is matter continuous or discontinuous in structure? Suppose this paper were divided into 10 pieces and then each piece subdivided into 10 smaller pieces and then each of these subdivided again; could such a process be repeated indefinitely producing smaller and smaller bits of paper or would we eventually get little bits that were no longer paper, or even little bits that could not be subdivided? In the last analysis, is all matter granular like sand or uniform as water appears to be? The question was never settled by the philosophers but has been settled by three centuries of experimental science. We know now that all matter is made up of very minute discrete particles called atoms and that there are only a comparatively small number (about 250) of kinds of atoms. In this chapter, we shall review briefly some of the evidence for this conclusion.

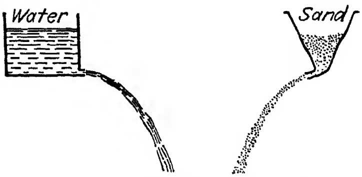


FIG. 1.—Continuous stream of water and discontinuous stream of sand.

3. It is always easier from the point of view both of the teacher and of the student to state a conclusion before giving the evidence on which it rests. In fact, there is always a temptation in discussing subjects too large for the time that is allotted to them to state conclusions without any consideration at all of the bases on which they rest. To teach physics in such a way is to miss the very essence of the subject. The ability to repeat Newton's three laws of motion does not constitute an understanding of the principles of mechanics, nor does an essay describing our present views about atoms and electrons give any real comprehension of modern physics. The concept of an atom is not one that is familiar in ordinary experience. We cannot have a proper understanding of this concept until we know something about the evidence that suggests the existence of atoms. Our concept will then be gradually built up and modified as it is used to correlate more and more observations. From the strictly logical point of view, therefore, we should not have made the statement in the last paragraph that "all matter is made up of very minute discrete particles called atoms, etc." We should rather have proceeded directly to a discussion of the evidence and shown what

conclusion could be drawn from it. As so often happens even in the development of a strictly logical subject like physics, the logical method is not the most practical. We shall frequently find it desirable to state a theoretical result before giving an account of the experiments that it is formulated to explain. Sometimes it will be necessary to omit a large part or all of the experimental data, though every effort will be made to avoid this. But the student is urged to reconsider every case where the experimental data are presented and endeavor to reconstruct the conclusions from the evidence.

4. Like most scientific evidence, that for the existence of atoms is indirect. None of our five senses give any indication of their presence. Certainly we see nothing in everyday life that would lead us to believe that everything is made of tiny particles less than a ten-millionth of a centimeter in diameter. Nor does the flow of billions of molecules of nitrogen and oxygen into a man's lungs with every breath give any sensation of the atomicity or discrete structure of matter. Our belief in such a structure is based on more complicated observations and reasoning. But, in recent years, a few experiments have been devised which show directly the effects of single atoms. These experiments were first performed long after the atomic theory was firmly established. If they had been done earlier, they might have remained unexplained or even have been explained satisfactorily without the idea of atoms.

5. These experiments depend on having a very few atoms going with enormous speeds, speeds so great that a single one of these atomic projectiles has energy enough to produce observable effects. It is not easy to produce such high speeds in the laboratory, but fortunately radioactive substances like radium furnish them. The so-called alpha particles ejected continually by many radioactive substances are very fast moving helium atoms (more strictly, as we shall see later, they are helium atoms that have lost two electrons). They move so rapidly that when they are stopped suddenly by striking a fluorescent substance they give up their energy in a splash of light, much as a stone hitting the surface of a pond gives a splash of water. These splashes of light, or scintillations as they are called, are not very bright, but they can be seen by using a low-power microscope in a darkened room. It is the accumulated effect of many of these flashes that

gives the luminosity to radiolite watch dials. In fact, a well-rested eye in a dark room can see the separate scintillations on such a dial by using an ordinary magnifying glass. Various apparatus and novelty companies sell little tubes called spinthariscopes which contain a speck of radioactive material, a fluorescent screen, and a small lens all mounted as a unit. Even to the unsophisticated, the twinkling flashes of colliding atoms in a spinthariscope are pretty things to see. Their fuller beauty can be seen only by the eye of the imagination trained to understand their significance and to appreciate the worlds that they reveal.

6. There are two other ways in which individual alpha particles can be observed. Alpha particles plowing their way through air break up the molecules of oxygen and nitrogen along their path into electrically charged pieces. This effect can be used to set up electrical currents which can be amplified and used to actuate a loud speaker or electrical recording device. An arrangement of this sort is called a Geiger counter or Geiger-Mueller counter. Another way of using the charged pieces of oxygen and nitrogen (called ions) produced by the alpha particles is the Wilson cloud chamber. In this apparatus, fog condenses on the ions marking the path of the alpha particle. Bright light shines on this line of fog and is reflected into a camera giving, in effect, a photograph of the path of the alpha particle. Such a photograph is shown in Fig. 2.

7. We have asserted that the effects described above are caused by individual atoms. By the end of this book, we hope that it will be clear whether or not this is a reasonable assertion. We have not proved the existence of atoms. We cannot prove it by any single experiment, but we can follow some of the lines of experiment and reasoning that have contributed to the proof. It is at least as important for the student to examine the method of establishing the conclusion as to learn the conclusion itself. New evidence might upset the conclusion, but the method of treating the new evidence would be essentially the same as the method of treating the old.

Matter

8. Since the first convincing evidence for the atomic structure of matter came from the study of chemistry, and since many of the technical terms which are useful in discussing problems

involving the structure of matter were originated by chemists, we shall devote the rest of this chapter to a brief survey of elementary chemistry.

9. Anyone who lives in the complex environment familiar to most of us encounters a great variety of material things. An Eskimo may need but two words to describe all the different liquids he knows, oil and water, but any car driver adds gasoline to his vocabulary and has a corresponding concept in his mind.

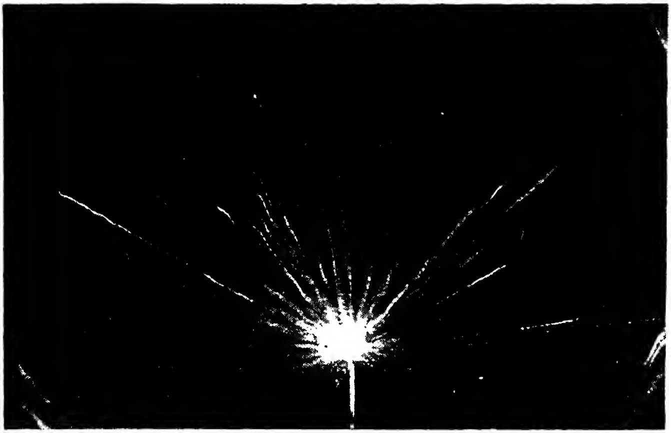


FIG. 2.—Tracks of alpha particles emerging from a radioactive source. (See also Chaps. XXI and XXVII.) (*Early photograph by C. T. R. Wilson.*)

Probably he is familiar with many other liquids such as alcohol, prestone, etc. He even knows that air is not the only kind of gas and, of course, he is familiar with an almost infinite variety of solid materials. It is with the differences and similarities between various materials and the way in which materials change that chemistry has to do. All materials are included in the comprehensive term "matter." All the stuff of which the material universe is made is called "matter" by chemists and physicists. Matter may be defined as anything that possesses weight or mass, anything that occupies space. It may be perceived ordinarily by the sense of touch and usually, though by no means always, by the sense of sight. As we have suggested already, matter exists in three different states, the solid, the liquid, and the gaseous.

The land, the sea, and the air are familiar examples of these three states.

Physical Properties

10. The many different kinds of matter with which we are familiar have different qualities and characteristics. These qualities and characteristics are known scientifically as "properties." Besides the qualities immediately apparent to the senses, such as color, smell, and feel, any given kind of matter has other properties determinable by various measurements, simple or complex. We know, for example, that alcohol and prestone freeze at lower temperatures than water. We can specify the properties of pure ethyl alcohol more accurately by saying that it is a colorless liquid freezing at −117.3°C. and boiling at 78.5°C. Or, to take a simpler substance that is less familiar but more useful in illustrating chemical problems, sulfur is a brittle yellow substance, rather light, nonmetallic, and with an acrid smell. So much we know from casual examination. By studying it more carefully, we find that it has a density of 2.06 g./cc. (*i.e.*, 1 cc. of it weighs 2.06 g.), it melts at 112.8°C., boils at 444.7°C., is insoluble in water, very soluble in carbon disulfide, is a nonconductor of electricity, and so on. We call such properties as these, *i.e.*, melting point, boiling point, solubility, electrical conductivity, etc., the physical properties of sulfur. We use them to give exactness to our concept of sulfur. If we found a piece of yellow material that looked like sulfur and was brittle like sulfur, we still would not be convinced that it was sulfur until we had made tests to determine its physical properties. The more properties we test and find to be identical with the set of properties that we have associated with the word "sulfur," the more we are inclined to call this new piece of material "sulfur" also. Obviously, there are practical limits to the number of tests we can make. Experience has shown us what properties or group of properties can be taken to be decisive. In practice, the chemist will have no difficulty in deciding whether a piece of pure material is sulfur or not, in spite of the philosophical difficulty of giving a completely exact definition of "sulfur."

Mixtures and Pure Substances

11. We have been careful to specify in the foregoing paragraph that we were dealing with "pure" alcohol or "pure" sulfur.

This was because there are many materials which have names and more or less definite characteristics that are not what we call pure substances; they are in fact mixtures of several substances, and their properties depend on the proportions of the mixture. For example, there are many different brands of gasoline on the market, all of which have the property of furnishing energy to an engine; but they are obviously not all the same substance. They have different colors, even their manufacturers admit that they have varying amounts of lead ethyl in them. They would not all have exactly the same boiling point, freezing point, ignition point, and so on. In other words, gasoline is not what chemists call a pure substance. Or again, the various hashes served in the local restaurants are obviously not pure substances. Sometimes it is possible to see that there are different materials in them, to separate the corned beef from the potato and the onion. At other times, the whole mixture is so finely divided and thoroughly intermingled that it appears to be a uniform conglomerate. But one would still hesitate to consider it a pure chemical substance and expect it always to show the same definite properties.

12. Most elementary chemistry texts attempt to set up an exact definition of a "substance," but such definitions are always more or less arbitrary, and such recent developments as the discovery of heavy water have emphasized the difficulty of making a definition that will hold universally. Fortunately, we can confine ourselves to cases of materials that are clearly mixtures of several substances or clearly pure substances, avoiding such dubious intermediate cases as metallic alloys or water.

Chemical Combination

13. As a typical example of a mixture, we may take finely divided sulfur and iron filings stirred up together. It is still possible to see with a magnifying glass the separate particles of sulfur and of iron. It is possible to separate out the iron by attracting it with a magnet or to dissolve out the sulfur with carbon disulfide. In other words, both the sulfur and the iron retain their identity and original properties. But if this mixture is put in a test tube and heated over a gas flame, it soon begins to glow with a dull red heat and when the test tube is allowed to cool a profound change is seen to have occurred. Instead of the previous greenish black powder, we now find in the test tube a black porous material which is entirely nonmagnetic and insol-

uble in carbon disulfide. It is an entirely new substance, quite homogeneous, and with properties of its own different from the properties of either iron or sulfur. We say that a chemical reaction has occurred and that a new substance called ferrous sulfide has been formed by the chemical combination of iron and sulfur.

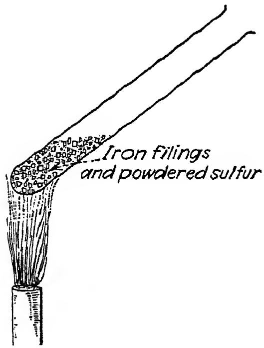


FIG. 3.—Iron-sulfur reaction.

14. There are, of course, many chemical reactions that are more familiar than that between sulfur and iron—the conversion of food into flesh and blood, for example, or the rusting of iron or the combustion of charcoal. The first of these reactions is much too complicated for us to consider, but rusting and combustion are combinations of simple substances with the oxygen in the atmosphere, differing from the iron-sulfur reaction only in the fact that one of the combining substances is initially in gaseous form.

15. These processes are changes that result in the production of new substances with different properties. Such changes are called chemical changes, and the processes that produce them are called chemical reactions. Other familiar examples which may be mentioned are the decay of animal and vegetable matter, the souring of milk, and the fermentation of grapejuice. There are other changes that can occur in matter changing some of its properties temporarily or in less fundamental ways which are called physical changes and are not considered to produce new substances. For example, color, smell, and feel were suggested above as qualities of substances that were directly perceptible, but a little consideration shows that they are not satisfactory properties by which to define a substance. Boiling water and ice water feel very differently yet they are certainly both water. One side of a piece of stone may feel rough and look white, whereas the other side of the same specimen may be a smooth polished blue. In other words, the same substance may have some of its qualities varying with differing conditions. We need not be overprecise in setting up a definition of what changes in properties are required before we say that one substance is

changed to another and what changes can be made without calling the altered material a new substance. Roughly speaking, changes resulting from melting, boiling, crushing, or polishing, that is, temperature treatment or mechanical treatment, are called physical changes and are not considered as producing new substances. On the other hand, the oxidation of iron to form rust or the burning of sulfur to form sulfur dioxide are said to produce entirely new substances having different densities, melting points, etc., from the parent substances. Such changes are called chemical changes.

Chemical Properties

16. Evidently not merely the melting point, boiling point, etc., are characteristic properties of a substance, but also its tendency to react chemically with other substances, whether it reacts with oxygen in rapid combustion, slow corrosion, or not at all, whether it combines with water or various acids, etc. These properties are called the chemical properties of a substance. Often it is easier to use them to determine the nature of a substance than to use the physical properties.

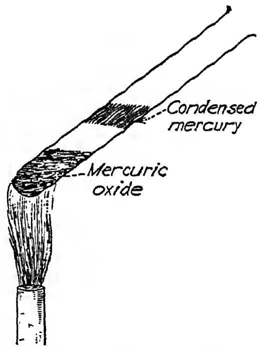


FIG. 4.—Decomposition of mercuric oxide.

Decomposition

17. There is a type of chemical reaction that we have not yet considered, namely, decomposition. This is the reverse of combination and consists in the breaking up of one substance into two or more other substances of entirely different properties. A good example of a simple decomposition is the breaking up of red oxide of mercury when it is heated. If some of this substance in powdered form is put into a test tube and held over a flame, gas is liberated which will cause a glowing splinter thrust into the test tube to burst into flame. More decisive tests show that this gas is oxygen. At the same time that the gas is being evolved, a mirrorlike deposit appears on the upper, cool part of the tube, and as the decomposition continues this deposit coagulates into droplets which are easily identified as mercury.

Elements and Compounds

18. The study of chemical reactions and, particularly, decompositions has led to very interesting conclusions about the nature of the various materials that exist in the universe. We have pointed out that their variety is almost infinite, but it was suggested long ago that all substances might be made up of a comparatively small number of elemental substances. The Greek philosophers thought that everything might be made of earth, air, fire, and water. We now believe again that all matter is built up from a small number (probably four) of different kinds of particles. But these are not the "elements" familiar to the chemist of the last hundred years. Later on, we shall discuss the probability that all the chemical elements can be broken into particles of which there are only four kinds, but at present we are concerned with the elements themselves. Their existence has been demonstrated by the study of such decompositions as that of mercuric oxide. Here we have a case of a brownish red substance not at all familiar to the average person which, on heating, breaks down into two comparatively familiar substances, oxygen and mercury. Or take the animal and vegetable world. It presents a great range of forms, colors, and structures. Yet the process of combustion shows that they all contain carbon as their chief constituent. It is found, in general, that there are only about ninety substances that cannot be decomposed. These substances are called elements. All other substances can be decomposed into two or more elements by some chemical process or other. Such complex substances are called compounds. We may sum up our definition of an element by quoting the words that Robert Boyle used in 1661. According to him,

"*The elements are the practical limits of chemical analysis, or are substances incapable of decomposition by any means with which we are at present acquainted.*"

19. Some of the substances familiar to us in daily life are elements, for example, iron, mercury, sulfur, carbon, oxygen, and nitrogen. Others like water, alcohol, or salt are compounds. Many others are mixtures, like air, which is chiefly oxygen and nitrogen but also contains some helium, argon, and other gases.

20. In the past forty years, a group of elements has been discovered which decompose spontaneously into other elements.

These elements are known as the radioactive elements, and the process by which they decompose is known as radioactive disintegration or merely radioactivity. In the past five years, it has been found possible to effect a similar decomposition of many other elements by artificial means. Nevertheless, the type of decomposition occurring in these cases is so *entirely different* from ordinary chemical processes that it seems best to continue to call these substances elements. When we return to the topics of natural and artificial disintegration later on in the book, we shall see that the 90 odd chemical "elements" are probably all built up out of three or four kinds of fundamental particles, but this in no way impairs the usefulness of the definition that we have given above.

21. We have good reason to believe that there are actually 92 different elements occurring in nature. Some of these are very common and have been known from the beginning of time. Others are extremely rare and have been discovered only recently. Two are entirely unknown, and the discovery of one or two others is as yet unsubstantiated. Many elements occur in nature only in compounds. Only about twenty occur in the free state, *i.e.*, not combined with anything else. Of these, we may mention gold, silver, platinum, copper, helium, hydrogen, oxygen, and nitrogen. The most abundant elements are oxygen, silicon, aluminum, iron, calcium, sodium, potassium, magnesium, hydrogen, titanium, chlorine, and carbon.

22. A list of all the known elements together with the abbreviations that are universally used to designate them is given in the table on the inside cover of this book. The other data given there will be explained from time to time throughout the book as occasion demands. In terms of the symbols given in this table, we would represent the iron-sulfur reaction as

$$Fe + S \rightarrow FeS.$$

Such a representation is called a chemical equation. This particular one simply says that it is possible to make iron and sulfur react chemically to give ferrous sulfide as a product. Like a mathematical equation, a chemical equation is merely a concise way of making a statement, a special kind of statement, to be sure, since it always describes a chemical reaction.

Conservation of Mass

23. Presently we shall consider several simple chemical reactions in detail with particular reference to the masses* of the substances involved. It was such studies that led to the atomic theory, but the path was long and indirect. The first step, a step necessary before a quantitative study of reactions could have any significance, was the proof of the law of conservation of mass. Stated in modern form, this Law of the Conservation of Mass is as follows:

The mass of a system is unaltered by any chemical change that occurs within it; or, when a chemical process is carried out, the total mass of the reactants is equal to the total mass of the products of reaction.

Or more briefly, matter can be neither created nor destroyed. This law is so familiar to us that we sometimes forget that it is not obvious but required many years of painstaking work for its establishment. Nor is it always recognized in the modern world outside of technical circles. Efforts to get something from nothing and claims of having done so still attract the gullible.

Law of Definite Proportions

24. In discussing the sulfur-iron reaction, we made no mention of the proportions in which the two substances were to be mixed but, as a matter of fact, the change we described is complete only for very definite proportions. If an excess of sulfur is used, some of it remains unaffected and the result of the reaction is a mixture of sulfur and ferrous sulfide. Similarly, if an excess of iron is used, some iron remains uncombined. Evidently a given quantity of sulfur reacts only with a given amount of iron, and vice versa. To be specific, if we start with 100 g. of iron, it will combine completely with 57.4 g. of sulfur, making 157.4 g. of ferrous sulfide, and the addition of more sulfur will not increase the amount of ferrous sulfide formed. Similarly, if we consider the burning of magnesium in oxygen, the reaction that gives the bright light used in flashlight photography, we find that for every gram of magnesium burned we get 1.66 g. of magnesium oxide.

25. Similarly, in the decomposition of any given substance, the proportion of the products of decomposition is always the same.

* The student will recall that the mass of a substance is proportional to its weight and is an expression of the amount of matter in it.

With mercuric oxide, for example, it is found that for every 100 g. of the oxide present initially there are always 92.61 g. of mercury and 7.39 g. of oxygen produced.

26. These results are typical of those obtained with all chemical reactions and are expressed in the important general law known as the Law of Definite Proportions.

A particular compound always contains the same elements chemically united in the same proportions by weight.

Law of Multiple Proportions

27. One of the surprising things that one learns in elementary chemistry is that not only is it possible for two elements to combine to give a substance with properties entirely different from either constituent element, but that two elements can combine in more than one way to form more than one compound. Carbon and oxygen, for example, can combine to form carbon monoxide, which is a deadly poison, or to form carbon dioxide, which is a harmless gas formed in the lungs in the process of breathing. Each of these gases is a pure stable substance with a definite freezing and boiling point and other specific properties. Each has a definite composition, and when we compare the compositions of the two we find an interesting result which we shall see presently has great significance. In carbon monoxide, the proportion of oxygen to carbon by weight is found to be always 16 to 12, while in carbon dioxide it is always 32 to 12, just twice as great.

28. A more complex case is presented by the combinations of nitrogen and oxygen. These two elements, though they exist side by side intimately mixed in the atmosphere without reacting at all, can under the right conditions combine in no less than five different ways to form five distinct compounds. The names of these five substances and their percentage composition by weight are given in the table on page 14.

In the next to the last column of this table, the number of grams of nitrogen corresponding to 100 g. of oxygen is computed to one decimal place for each compound. It can be seen that the second amount is one-half the first, the third amount one-third the first, the fourth one-fourth the first, and the fifth one-fifth the first. Similarly, when we examine the last column which gives the amount of oxygen per 100 g. of nitrogen for each

compound, we see that the numbers are approximately 57, 57 × 2, 57 × 3, 57 × 4, and 57 × 5.

Name	Percentage nitrogen	Percentage oxygen	Grams of nitrogen per 100 g. of oxygen	Grams of oxygen per 100 g. of nitrogen
(1) Nitrous oxide.........	63.65	36.35	175.1	57.1
(2) Nitric oxide...........	46.68	53.32	87.5	114.3
(3) Nitrogen trioxide......	36.86	63.14	58.4	171.3
(4) Nitrogen tetroxide.....	30.45	69.55	43.8	228.4
(5) Nitrogen pentoxide....	25.94	74.06	35.0	285.5

29. Examples of this sort might be multiplied indefinitely, particularly by citing the innumerable carbon-hydrogen compounds. The results may all be summed up in the following law, called the Law of Multiple Proportions:

When two elements A and B unite to form more than one compound, the weights of B that unite with a fixed weight of A stand to each other in the ratio of whole numbers which are usually small, and vice versa.

30. These fundamental chemical laws, the law of definite proportions and the law of multiple proportions, are now to be considered in the light of the questions raised earlier about the structure of matter. The very existence of chemical reactions suggests that matter is not uniform, cannot be indefinitely subdivided and maintain its nature. The laws that govern the reactions point strongly to the atomicity of matter. Take the example discussed above, the oxides of nitrogen. Let us consider how readily the existence of these compounds is explained in terms of atoms. First we shall make an incorrect assumption which nevertheless offers a simple explanation of the data we have given. Then we shall see that a modification of our first assumption is equally satisfactory and is in fact the correct explanation. Assume then that nitrogen is made up of indivisible atoms each weighing 10 x millionths of a gram and that oxygen is made up of indivisible atoms each weighing 5.71 x millionths of a gram where x is unknown. Then the simplest compound of oxygen and nitrogen that can be imagined is one where every atom of nitrogen is attached to one and only one atom of oxygen.

Such a compound would be made up of thousands of particles each containing one atom of nitrogen and one of oxygen. The proportions by weight of nitrogen to oxygen in any sample of such a compound would be simply the ratio of the weights of the separate atoms. On the foregoing assumption, the ratio of oxygen to nitrogen in this simplest possible compound would be 5.71 :10.00 which is exactly the ratio found from the last column of the table of Par. 28, page 14, for the first compound, nitrous oxide. Other compounds might exist where each unit particle or molecule,* as we call it, contained several atoms of oxygen to each atom of nitrogen, or vice versa. Applying this idea to the data of composition by weight given in the table, we would conclude that the molecules of each of the five compounds contained one atom of nitrogen but that the number of oxygen atoms in each molecule increased from one to five as we went down the list. Thus the atomic hypothesis has given a simple and satisfactory explanation of the data given. But it may not be the only explanation, and perhaps it cannot be reconciled with other data which we have not presented.

31. Suppose we try again using a somewhat different assumption. Perhaps the atom of oxygen is a little heavier than that of nitrogen instead of being half as heavy. If we assume it to weigh $11.43x$ millionths of a gram while still keeping the weight of a nitrogen atom as $10x$ millionths of a gram, evidently it must be the second of the nitrogen-oxygen compounds that is the simple one-to-one structure. The first is still perfectly understandable. If it contains two atoms of nitrogen and one of oxygen, its percentage composition will be as shown. Similarly, the compositions of the three other compounds are explicable in terms of the second assumption about the weights of the atoms. This is clear from the chemical formulas for all five compounds as given below:

(1)	(2)	(3)	(4)	(5)
N_2O	NO	N_2O_3	NO_2 or N_2O_4	N_2O_5

where the symbols N and O stand for nitrogen and oxygen (see the chemical tables in the back of this book or any list of elements

* We use the word molecule for the smallest separate particle of a substance. It will contain two or more atoms if the substance is a compound but is also sometimes used for elements where there may be only one atom in each molecule.

and their symbols) and the subscripts indicate the number of atoms in the molecule. Evidence from other chemical reactions shows that the second assumption and the corresponding formulas above are correct.

32. If we examine the arguments of the last paragraph, it is evident that the conclusions drawn about the number of nitrogen and oxygen atoms in the molecules of the various compounds did not depend at all on the actual values assumed for the weights of the nitrogen and oxygen atoms but only on their relative values. The chemical formulas given would follow equally well from a choice of 10 and 11.43 mg., 20 and 22.86, 14 and 16, or any other values for the respective weights of the atoms of nitrogen and oxygen so long as the ratio was the same. Conversely, it is clear that the percentage composition by weight of the five nitrogen-oxygen compounds is not in itself sufficient to determine either the relative or absolute mass of the atoms of oxygen and nitrogen. It does, however, suggest very strongly the existence of such atoms. The absolute masses of atoms are determined by entirely different methods described in Chap. XX. Their relative masses as expressed by the atomic weights (see below) can be determined only by the comparison of the analyses of many compounds and the study of many reactions involving various elements. Such a further study shows that it is possible to find a system of combining weights that can be used for all compounds, a result that follows naturally from the atomic hypothesis but that would be difficult to explain otherwise. For example, consider again the elements hydrogen, carbon, nitrogen, and oxygen. In water (H_2O), we have the proportions by weight of hydrogen to oxygen 2 to 16, in hydrogen peroxide (H_2O_2) it is 2 to 32. We saw that carbon and oxygen combined in the ratios of 12 to 16 or 12 to 32. In the nitrogen-oxygen compounds, the ratios of oxygen to nitrogen are

(1)	(2)	(3)	(4)	(5)
N_2O	NO	N_2O_3	NO_2	N_2O_5
$\frac{14 \times 2}{16}$	$\frac{14}{16}$	$\frac{14 \times 2}{16 \times 3}$	$\frac{14}{16 \times 2}$	$\frac{14 \times 2}{16 \times 5}$

In the hydrogen-carbon compounds, the ratios of hydrogen to carbon are $\frac{1}{12}$, $\frac{1 \times 2}{12}$, $\frac{1 \times 3}{12}$, $\frac{1 \times 4}{12}$, etc., and the familiar nitrogen-hydrogen compound ammonia gas (NH_3) has 14 g. of

nitrogen for every 3×1 g. of hydrogen. We see therefore that the compounds of the four elements oxygen, hydrogen, carbon, and nitrogen are all consistently interpreted if we assign the weights 16, 1, 12, and 14, respectively, to the elements in question.

33. These examples of chemical reactions need not be multiplied further; a year's course in chemistry does not exhaust those that are known. They are all more or less simply explained by the following hypotheses of the atomic theory:

1. Every element is made up of exceedingly small, indivisible particles called atoms.
2. The average weight of the atoms of a particular element is always the same but is different for different elements.
3. Chemical compounds are formed by the union of the atoms of different elements in the simplest numerical proportions.

The Atomic Weight

34. From the study of chemical reactions, no very good estimate of the size of an atom can be made. Since reactions of the most minute quantities that can be studied behave just like the reactions between large quantities of material, it is certain that the atoms are much smaller than the fractions of milligrams that can be handled by a chemical balance. Nor can they be seen with a microscope. They are small, therefore, but their exact size must be determined by other means. But though we do not know their absolute size and weight, we do know their relative weight. In the carbon-oxygen compounds cited above, for example, the carbon monoxide molecule (CO) has one atom of carbon and one of oxygen, the carbon dioxide molecule (CO_2) has one atom of carbon and two of oxygen; in each case, quantitative analysis shows the weights of the carbon and oxygen atoms are in the ratio 3:4. Similarly, by studying large numbers of compounds, it has been possible to establish the relative weights of all the known elements. These might be expressed on any scale. The weight of an atom of oxygen might be called 100, and on this scale the weight of an atom of carbon would be 75. This would not mean that a carbon atom weighs 75 lb. or 75 g. or even 75 millionths of a gram. It would mean that the average weight of the carbon atoms studied in chemical reactions was $\frac{75}{100}$ of the average weight of the oxygen atoms studied in chemical reactions. In practice, it is found most convenient to assign the

number 16.000 to the atoms of oxygen and express ther elative average weights of the atoms of the other elements with reference to this. The numbers obtained in this way are called the atomic weights of the elements. These atomic weights are given in the periodic table of the elements found in the back of this book. A few of the common elements are given here.

Element	Symbol	Atomic weight
Oxygen...	O	16.000
Copper....	Cu	63.57
Hydrogen.	H	1.008
Aluminum	Al	26.97
Sodium...	Na	22.997
Chlorine..	Cl	35.457
Carbon....	C	12.010

35. In the case of compound substances where each molecule contains more than one atom, the molecular weight is defined as the sum of the atomic weights of the atoms in the molecule. For example, the approximate molecular weights of the five nitrogen-oxygen compounds are $N_2O = 44$, $NO = 30$, $N_2O_3 = 76$, $NO_2 = 46$, and $N_2O_5 = 108$.

36. It should be emphasized that the atomic weight is determined by chemical experiments each of which involves many thousands of atoms. The constancy of the atomic weight of a particular element as determined by different investigators does not prove that all the atoms of that element have the same weight but only that the average weight of a large number is the same. The importance of this point will be seen later on when isotopes are studied.

37. According to atomic theory, the weights of equal numbers of atoms of two elements should be in the ratio of the atomic weights and vice versa. For example, 12 lb. of carbon should contain the same number of atoms as 16 lb. of oxygen. It is found convenient to give a name to that particular quantity of an element which has a weight in grams numerically equal to the atomic weight. Such an amount of any element is called a gram atom or mole of that element. Similarly, *for any compound substance the quantity of the substance that has a weight in grams numerically equal to the molecular weight of the substance is called*

a gram molecule or mole of the substance. For example, a mole of carbon dioxide is 44 g. In a later chapter, the determination of the number of molecules in a mole will be discussed. This number is evidently the same for all pure substances whether elements or compounds. It is called Avogadro's number and has the value 6.03×10^{23} molecules per mole.

38. The existence of atoms forms a common ground from which the study of both modern chemistry and modern physics must start. The discovery of the 92 different elements, their reactions with each other, and their arrangement in the periodic table have been largely the domain of chemistry. It is only quite recently that physics has concerned itself much with the differences between different elements and the atomic structure which causes these differences. Later on in the book, we shall return to these problems. At present, we shall discuss the behavior of atoms and molecules without very much stress on the differences between the atoms of different elements.

SUMMARY

The interlocking character of physics and chemistry is discussed. The idea of the atomic nature of matter is introduced and some simple experiments described which show the effects of individual atoms. After explaining what is meant by mixtures and pure substances, by physical and chemical properties, a few typical chemical reactions are described in detail. The difference between elements and compounds is explained. Three laws are then stated which sum up the results of the study of chemical reactions. They are

The Law of Conservation of Mass
The Law of Constant Proportions
The Law of Multiple Proportions.

These laws are then interpreted in terms of the atomic hypothesis. Atomic weights are defined. The mole or gram molecule is defined and Avogadro's number given.

PROBLEMS

1. What are the physical properties of a substance?

2. How is it possible to distinguish a pure substance from a mixture?

3. Which of the following materials are mixtures, which are chemical compounds, and which are elements?

Wood	Charcoal	Glycerin	Glass
Paper	Diamond	Table salt	Wine
Tar	Brass	Rubber	Rust
Coal	Aluminum	Graphite	Flashlight powder
Carbon	Arsenic	Lead	Ether
Air	Chromium	Water	Silver
Carbona	German silver	Nickel	Baking soda
Laughing gas	Mercury	Sand	Sugar
Cyanogen	Steel	Neon	Solder
Manganese	Ice	Helium	Earth
Copper sulfate	Zinc	Potatoes	Radium

4. Write two sets of formulas for the compounds of lead and oxygen based on the following table of weights of the compounds, and state which one is correct.

Name	Grams of lead per 100 g. of oxygen	Grams of oxygen per 100 g. of lead
Lead monoxide	1,294	7.72
Lead dioxide	647	15.44
Lead red oxide	972	10.29
Lead sesquioxide	863	11.58
Lead suboxide	2,590	3.86

5. How many atoms are there in 1 g. of oxygen, 1 g. of aluminum?

6. Assume that air is one-fifth oxygen by weight and has a density of 1.29 kg./cu. m. If all the oxygen in a sealed room is converted to carbon dioxide, what is the density of the gas in the room? If the volume of the room is 30 cu. m., how many grams of carbon dioxide have been made?

7. How many pounds of air are required to burn 20 lb. of charcoal if half the charcoal burns to carbon monoxide and half to carbon dioxide?

8. In the oxides of copper, compute the number of grams of copper per 100 g. of oxygen.

Copper suboxide	Cu_4O
Cupric oxide	CuO
Cuprous oxide	Cu_2O

9. Two elements A (atomic weight 60) and B form three compounds. The simplest AB contains 60 per cent by weight of A. The other two compounds contain 20 per cent and 80 per cent by weight of A, respectively. What is the atomic weight of B, the composition of the two unknown compounds, and the molecular weight of all three?

10. If 10^{25} atoms of sodium weigh 380 g., how much do 10^{25} atoms of chlorine weigh?

11. How much does one atom of oxygen weigh, one atom of hydrogen?

12. How many grams of carbon will contain the same number of atoms as 100 g. of carbon dioxide (CO_2)?

13. A mixture of 80 g. of sulfur and 80 g. of iron filings are mixed and heated until they react chemically. How much ferrous sulfide is formed, and how much sulfur is left over?

14. The weight of the hydrogen in the airship *Hindenburg* was about 16,000 kg. How much water did it form when it burned?

15. A compound known to contain only carbon and hydrogen has a molecular weight of 30, what is its most probable chemical formula? What other formula satisfies this condition?

16. What is the percentage composition by weight of sodium chloride ($NaCl$)?

17. State the evidence for the existence of molecules.

CHAPTER II

MOLECULAR SIZES, MOTIONS AND FORCES

1. In the first chapter, we have presented evidence for the discrete structure of matter and the existence of atoms. It has been suggested that all matter is made of some 92 different elements. On this theory, if any bit of matter, gaseous, liquid, or solid, could be sufficiently magnified, it would be seen to consist of tiny particles and each of these particles could be identified as belonging to one of the 92 known elements. Actually, such a magnification is impossible, a fact which in itself has interesting philosophical and scientific implications but which for the moment we accept without further consideration. Therefore the correctness of our theory must be tested by some more indirect method.

2. The method used is typical of scientific procedure in general. The hypotheses suggested by the facts given in the first chapter are amplified and applied to the interpretation of other phenomena. Since they are successful in interpreting well-known physical processes and in predicting others previously unknown, the theory is said to be true.

3. The development of the atomic theory of matter and the interpretation of various physical phenomena in its terms was one of the great intellectual triumphs of the nineteenth century. Historically and logically, the proper way to describe this development would be to start with a factual description of various phenomena and then proceed to their interpretation, building up the theory gradually. But it is simpler to assume the general results at the start and present some of the relevant phenomena along with their interpretation.

4. According to our present views, all substances are made of a very large number of particles submicroscopic in size called atoms. There are forces of attraction between these atoms which vary greatly from one element to another and cause the formation of chemical compounds. This much we have already discussed. But these forces of attraction are not satisfied by forming com-

pounds. When atoms combine to form closely knit molecules, the atoms in one molecule may still have some attraction for the atoms in neighboring molecules. In fact, in crystalline solids the mutual attractions between the atoms are so complex that it is impossible to distinguish one molecule from another although, for our purposes, we can think of any compound or element as made up of clearly defined units containing one or more atoms, and we may consider these units as indissoluble by physical means. There will be forces of attraction between these molecules, and there must also be forces of repulsion that become important when the molecules come close enough to each other. These ideas which arise from those of the last chapter are not quite enough; we now add that of the motion of the molecules, a motion chaotic in character but increasing with temperature. Let us recapitulate by writing down our assumptions explicitly in tabular form as follows:

1. Every pure substance is made up of a large number of similar minute particles called molecules.

2. There are forces of attraction between these molecules which decrease very rapidly as the distance between molecules increases.

3. There are forces of repulsion between the molecules which decrease even more rapidly with distance than the attractive forces. In fact, these forces are negligible except when the molecules are practically "in contact."

4. The strengths of the intermolecular forces as well as the sizes, shapes, and masses of the molecules depend on the nature of the substance.

5. The molecules are constantly in motion, colliding with each other, and the average amount of this motion increases as the temperature rises. At all temperatures, some molecules are moving very rapidly and some very slowly.

5. In the case of some substances, the forces of attraction between the molecules are so large that at ordinary temperatures their motion is restricted to one of oscillation about a fixed position. The molecules are closely packed and so strongly held together that they form a rigid structure. Such a substance is called a solid. Speaking technically, it is a substance that offers resistance to a shearing force. A solid body has a definite shape which it maintains in spite of the gravitational forces tending to level it down. In fact, it usually requires very considerable

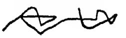
forces to change the shape of a solid. All known substances can be solidified by reducing the temperature sufficiently, though in the case of helium the molecules must be almost completely deprived of motion to achieve this, so small is the force of attraction between them.

6. In other cases at ordinary temperatures, the forces of attraction between the molecules are not sufficient to keep them in fixed positions but do suffice to keep them close to each other. They may be thought of as sliding over each other, moving in and out, but always remaining closely packed, as much so as in a solid. Such a substance loses its rigidity. It is a liquid, lacking resistance to shearing stress and therefore flowing more or less readily under the action of gravitational or other forces. In contrast to a gas, a liquid has a free surface.

7. Finally, if the motion of the molecules is great enough, they are forced apart by mutual collisions, the mutual attractions at greater distances become rapidly smaller, and therefore the tendency for the molecules to cling together is almost entirely lost. They fly apart as far as they can. In this case, we have a gas, a state characterized by the fact that the substance tends to fill completely any vessel into which it flows.

8. So much for the general assumptions of the kinetic theory of matter. There are three main points: the existence of molecules, their motion, and the forces between them. The first point was established in the first chapter and will be reinforced from time to time later on. Of two striking proofs of the second point to be presented, the Brownian movement has been known for a long time, the direct measurement of molecular velocities is very recent.

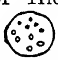
The Brownian Movement

9. According to the views of the atomic theory, the motion of molecules in a gas or liquid is extremely chaotic. A molecule may move a certain distance in one direction, strike another molecule, and bounce off at an angle, moving on to another collision. Its path for a fraction of a second might look like this . Now, as we said, a molecule is submicroscopic. It cannot be observed. But particles of microscopic size, though containing many thousands of molecules, are yet so small that they partake somewhat of this chaotic motion. Thus, a

smoke particle in air may be struck by many hundreds of air molecules in a hundredth of a second. Yet it may be struck on one side by more hundreds than on another and may get a resultant motion in a particular direction. Then in the next hundredth of a second the preponderance of collisions may be from another direction, giving a different resultant motion so that the path of the particle may look like this . As particles grow larger, the probability of their being hit harder on one side than on the other decreases rapidly—but it does remain finite. The table on which I write might rise up and slap me in the face if, for an instant, no molecules hit the top and many billions hit the bottom.

10. The movements described above were first observed by a botanist, Robert Brown, in 1827, long before their explanation was understood. Since that time, they have not only been explained qualitatively, but the details of the motion have been measured and quantitatively explained by the mathematical kinetic theory of gases. Under a low-power microscope, they are easily observed in tobacco smoke and can also be seen in fine particles of gamboge or India ink in water.

Diffusion

11. The particles observed in the chaotic motion known as the Brownian movement may be made smaller and smaller. As they diminish in size, their motions grow more and more violent, more and more like the motion of the molecules colliding with them. Suppose that 10 particles are under observation in the field of the microscope. Suppose that five of them are of gamboge and five of India ink and that to begin with they are separated thus . Then, if they are watched for a few moments, it will be seen that their random motion under the impacts of the molecules of the liquid causes them to become mixed thus . This same process takes place among the molecules of a gas or liquid and is called diffusion. Its occurrence is most frequently noticed in ordinary experience by our sense of smell, which is so sensitive that it can detect an extremely small concentration of a smelly substance in the air, as little as one part in a billion under favorable circumstances. Thus it is not surprising that when a bottle of ammonia, ether, hydrogen sulphide,

or any other volatile substance having a strong smell is opened in a room, the whole room is gradually filled with the reek of it, even if the air of the room is perfectly still, free from convection currents of all kinds. This results from the diffusion of the odoriferous molecules from the vapor in the bottle into the air of the room. The simultaneous diffusion of the air molecules back into the bottle is a phenomenon which is less easy to detect but which can be shown to occur.

Speed of Diffusion of Different Gases

12. From the qualitative observations of the kind mentioned in the last paragraph, we can determine that molecules diffuse through the atmosphere remarkably rapidly. Even on a perfectly still night, it does not take long to learn that a skunk has passed. We shall not concern ourselves with exact quantitative determinations of rates of diffusion. There is one very simple experiment, however, which we shall describe since its explanation is very suggestive.

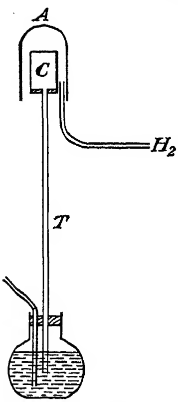


FIG. 5.—Diffusion of hydrogen.

13. Suppose we take a porous earthenware vessel such as is still used in hot dry countries for cooling liquids. For convenience, it can be in the shape of a cylindrical cup. Close the cup with a stopper through which runs a piece of glass tubing. Mount the cup so that the tube comes down vertically from it into a bottle partly filled with water as shown in Fig. 5. Now put an inverted beaker over the porous cup and run some hydrogen into it through the rubber tube as shown. Hydrogen is lighter than air and will tend to displace the air in the beaker, surrounding the porous cup with hydrogen. The molecules of air inside the cup will gradually diffuse out through its porous walls and the hydrogen molecules outside will tend to diffuse inward. If the two kinds of molecules diffuse at the same rate, as many will go in as come out; but if the rates of diffusion are different, the number of molecules inside the cup will tend to increase or diminish. Actually we observe that as soon as the hydrogen starts to flow into the space around the cup, water is forced down in the tube into the wash bottle until gas is actually bubbling up through the water and forcing a stream out the

nozzle. This continues only for a short time. The bubbling and water jet gradually die out, indicating that a condition of equilibrium has been established. If now the beaker is taken away so that the cup is once more surrounded by air, water is sucked up from the bottle into the cup. Since this is undesirable, the experiment is usually discontinued before the water has risen so far. This complicated chain of events is easily explicable on the assumption that hydrogen diffuses more rapidly than air. At first, the cup is full of air and there is hydrogen outside. The hydrogen comes in faster than the air goes out, thereby increasing the number of molecules in the cup and forcing some of them down the tube. This continues until most of the air in the cup has had time to diffuse out and be lost in the atmosphere of hydrogen maintained in the beaker. When this has occurred, there is hydrogen both inside and outside the cup, so that there are just as many molecules diffusing out as in and there is no further building up of pressure inside the cup. When the beaker is taken away, the gas inside the cup is still hydrogen, but outside it is air once more and the hydrogen molecules find their way out faster than the air molecules find their way in. Consequently, the amount of gas inside the cup diminishes and the water is sucked up to take its place. Eventually the hydrogen would all diffuse out and its place be taken by air. Equilibrium would then be reestablished.

14. The preceding explanation was based on the assumption that hydrogen diffuses more rapidly than air. We could repeat the experiment with various combinations of gases. In this way, we could accumulate quite a collection of observed facts about porous cups and the rise and fall of water in tubes connected with them. It is conceivable that some of these observations might be of immediate usefulness, might be applied at once to the solution of some problem, but it is unlikely. However, when we find that we can express the result of all these experiments in the statement that light gases always diffuse more rapidly than heavy gases, we realize that we are describing a great many events in a single statement. We say that we have discovered a scientific truth. It is this reduction of many observations to a single one, this relation of many events to a single guiding principle that is the object of the scientific method. The mere collection of facts is not science.

15. So far, we have said merely that light gases diffuse more rapidly than heavy ones. Relating this observation to others that have suggested the existence of molecules, we say that light molecules move with higher average speeds than heavy ones (as long as the temperature is the same). Later on in Chap. VIII, we shall see that this statement can be made more precise and inclusive by the use of mathematical language, but we are not ready for that yet.

Diffusion in Liquids and Solids

16. In the general assumptions of the kinetic theory stated in Par. 4 of this chapter, motion was attributed not only to the molecules of a gas but also, in lesser degree, to those of liquids and solids. Diffusion might be anticipated, therefore, in liquids and solids also. This is found to be true, although, as might be expected, the process occurs much more slowly than in gases. For example, the water in an open pitcher standing in a refrigerator or sometimes the ice cubes themselves acquire the flavor of sundry foodstuffs occupying the refrigerator at the same time. This case involves diffusion through the air of the refrigerator and then into the water. A less domestic example is observed by putting a crystal of copper sulfate in a jar of water and watching the gradual spread of blue color throughout the whole volume of the water. The time needed for these processes is many hours in contrast to the few minutes needed for the analogous gaseous diffusion processes.

17. In discussing solids, it was stated that the molecules vibrated around positions of equilibrium and did not migrate from place to place throughout the substance. This is true in general, but we have pointed out that there are always some molecules having velocities much greater than average. Such molecules may occasionally break away from the forces holding them in fixed positions and migrate short distances through the solid. Evidence for such diffusion of the molecules of solids was found some years ago when a block of gold and a block of lead were clamped firmly together and left in contact for a period of many months. After separating them, thin slices cut parallel to the plane of contact were examined and it was found that each metal had diffused somewhat into the other.

The Size of Molecules

18. Molecules, we have said, are too small to be seen with a microscope. That means they must be less than $\frac{1}{1000}$ mm. in diameter, as this is about the limit of the best compound microscope. But how much smaller are they? There are several ways of guessing maximum possible diameters. For instance, a tiny drop of oil, say 1 cu. mm. in volume, will spread out over a surface of clean water giving a uniform film of some 30 by 30 cm. The film is therefore about one-millionth of a centimeter, *i.e.*, 10^{-6} cm., thick. It cannot be less than one molecule thick, and therefore a molecule of oil cannot be more than 10^{-6} cm. in diameter. Similarly, a small quantity of soapsuds can be blown into such a large bubble that the thinnest part of the bubble is not more than 6×10^{-7} cm. thick. Very thin but still uniform gold leaf on the other hand is 10^{-5} cm. thick, which pushes the possible diameter of a molecule to a much smaller figure than the microscope but one ten times larger than that allowed by the dimensions of an oil film.

19. More accurate data from more complicated experiments give values of 2 to 5×10^{-8} cm. for the diameters of common molecules containing two or three atoms. For oxygen (O_2), the diameter is 3.4×10^{-8} cm. The meaning of the "diameter" of something so small it can be neither seen nor touched is questionable. In terms of intermolecular forces, the diameter of a molecule can be defined as the distance separating the centers of two similar molecules which are so near that the repulsive force between them prevents any closer approach. The analogy of two colliding billiard balls is useful if it is not carried too far. For the present, a molecule may be considered a sphere occupying a space about 10^{-8} cm. in diameter so completely that no other molecule can share any part of it.

20. The masses of atoms can be determined in a number of ways, notably by the mass spectrograph which we shall discuss in Chap. XX. One of the simplest methods is to use the value of Avogadro's number which we have already quoted at the end of Chap. I. If we recall that this is the number of molecules in a mole and that the mass of a mole is the molecular weight, it is evident that the mass of a single molecule is the molecular weight divided by Avogadro's number. For hydrogen, such a

calculation gives $1.008/6.03 \times 10^{23} = 1.67 \times 10^{-24}$ g. for the mass of a single atom. The masses of the atoms of the other elements will be proportionately higher in the ratio of their atomic weights.

Space between Molecules. Mean Free Paths

21. The Brownian movements and the facts of diffusion may be accepted as establishing the general idea of molecular motion, but we have not as yet said how far molecules move between collisions or how fast they travel. To get some idea of the spacing of the molecules in a gas, we shall quote another experimental result which will be discussed more fully in a later chapter (VII), namely, that one mole of any gas at normal temperature and pressure (NTP = 0°C. and 760 mm.) occupies a volume of 22.4 l. From this, the number of molecules in 1 cc. can be calculated by dividing Avogadro's number by the number of cubic centimeters in 22.4 l. This gives

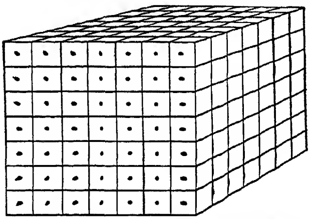


FIG. 6.—The average spacing of the molecules in a gas.

$$\frac{6.03 \times 10^{23}}{22.4 \times 10^{3}} = 2.7 \times 10^{19}$$

molecules per cubic centimeter at NTP. If the molecules were uniformly distributed and motionless, 1 cc. of gas might be represented by Fig. 6 where each dot is a molecule. In order to get a total of 2.7×10^{19} molecules in the cubic centimeter, there would have to be 2.7×10^{19} little cubic cells each containing one molecule. This means that along each edge there would be

$$\sqrt[3]{2.7 \times 10^{19}} = 3 \times 10^{6}$$

cells. Since the length of each edge is 1 cm., this means that the center of each molecule would be

$$\frac{1}{3.0 \times 10^{6}} = 3.3 \times 10^{-7} \text{ cm.}$$

from that of its nearest neighbor. Thus for oxygen, whose diameter has been given as 3.4×10^{-8} cm., this separation is almost exactly 10 diameters. This shows that there is a good deal of empty space in a gas, a fact which is confirmed by the

following simple calculation. The volume occupied by each molecule is

$$\frac{\pi d^3}{6} = 2.0 \times 10^{-23} \text{ cc.}$$

for oxygen. Therefore, $2.0 \times 10^{-23} \times 2.7 \times 10^{19} = 5.4 \times 10^{-4}$ cc. is the total volume occupied by the molecules themselves in 1 cc. of oxygen at atmospheric pressure and freezing temperature. In other words, all but about 1/2,000 of a room full of air is empty space. Diffusion and Brownian motion show that the molecules must be moving about pretty rapidly. It is also well established that, on the average, they move very short distances between collisions. In fact this distance, called the mean free path, is, for oxygen, almost exactly 10^{-5} cm., or 1/10,000 mm. at NTP. Perhaps it should not be called a small distance as it is nearly a thousand molecular diameters and therefore considerable from the point of view of a molecule. If the pressure of gas in a closed space is reduced by pumping out some of the molecules, then the chances for collisions are reduced and the mean free path correspondingly increased. It is in fact inversely proportional to the pressure if the temperature is kept constant. Modern vacuum technique makes it possible to remove so large a proportion of the gas from a vessel that the mean free path becomes as much as a meter or more. Even under poor conditions, a good mechanical pump can produce a vacuum in which the pressure is 0.1 mm., or about 10^{-4} of atmospheric pressure, so that the mean free path becomes ten thousand times greater than at ordinary pressure, or about one millimeter. "Hyvac" pumps are guaranteed to produce a pressure below 3×10^{-4} mm. of mercury under good conditions. If a mercury diffusion pump is also used the pressure may be reduced below 10^{-5} mm. or about 10^{-8} of atmospheric pressure. At this pressure, the mean free path is about 10 m.

22. The following experiment, illustrated in Fig. 7, (*a*) and (*b*), shows the dependence of the mean free path on pressure and the fact that molecules do move and move in straight lines if collisions can be avoided. An arrangement of two glass bulbs connected by a short tube is filled with air at atmospheric pressure. A little iodine is put in the bottom of the lower bulb. Iodine is solid at ordinary temperatures, and the molecules are held

together so strongly that they only occasionally break away, but if it is heated the molecules vibrate more vigorously and a great many of them break away. Immediately on leaving the surface, they collide with the air molecules and pursue random courses traveling an average of 1/10,000 mm. between collisions. Some of them eventually reach the cold glass walls where they condense. Some return to the solid iodine again. Some eventually diffuse up into the second vessel and after many thousand more collisions they may condense on the walls. If the heating of the iodine in the lower vessel is continued for a long time, a

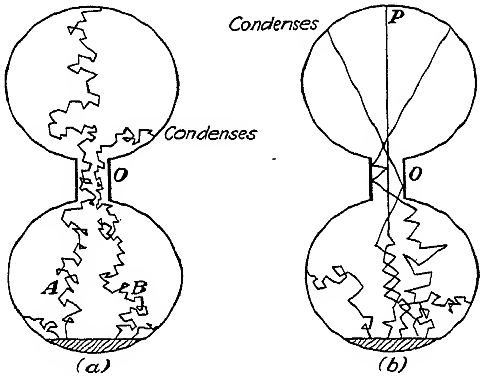


FIG. 7.—(*a*) The diffusion of iodine molecules in air at atmospheric pressure. (*b*) The diffusion of iodine molecules at reduced pressure.

thin deposit of iodine will condense all over the cold surfaces of the upper vessel. But it will be uniform. The molecules coming into the upper vessel will follow no definite course. They will bump drunkenly from one air molecule to another and are just as likely to end on one part of the glass as the other.

23. Now consider what happens if the air is removed. It is possible to do this so completely that the mean free path of the iodine molecules in the upper vessel is much greater than the dimensions of the bulbs. The iodine molecules themselves will be so numerous in the lower vessel that this may not be true there but that is immaterial. Under these conditions, very few collisions occur in the upper vessel, except with the walls, and molecules striking them stick. Therefore the only molecules that will get into the uppermost vessel will be ones that came

through the connecting tube, O, but such molecules all keep going in straight lines striking the upper part of the bulb. Therefore there should be a deposit of iodine in this region but not elsewhere. This is the observed result and is therefore in support of the hypotheses on which the prediction was based.

The Direct Determination of Molecular Velocities

24. The type of experiment we are about to describe would surely have brought joy to the hearts of such men as Clerk-Maxwell and Boltzmann who gave form to the mathematical theory of gases at a time when talk of molecules and molecular velocities seemed almost as abstract and unreal as does talk of four-dimensional space today. Unfortunately, no such direct proof of their ideas was possible then. It had to wait for the development of high-vacuum technique, to wait through a period of fifty years during which their theory had been so often proved and tested indirectly as to make more direct evidence hardly necessary.

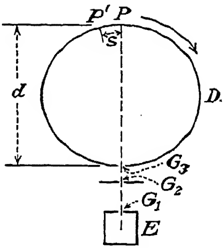


FIG. 8.—Zartman's apparatus for measuring molecular velocities.

25. Beginning with Stern at Hamburg in 1920, several experimenters have measured molecular velocities directly. The experiment most suitable for presentation here was done by Zartman at the University of California in 1930. A schematic diagram of his apparatus is shown in Fig. 8. E was an electrically heated furnace in which some metal could be vaporized. G_1 was a narrow slit in the side of the furnace, G_2 another slit held in position beyond G_1. A third slit G_3 was cut in the side of a cylindrical steel drum D which could be rotated at a high speed about an axis perpendicular to the paper. The whole system was in a box which was highly evacuated. Then E played the role of the lowest bulb in the iodine experiment described above, and the slits G_1 and G_2 corresponded to the tube O. Metal vapor molecules got to G_2 only if they emerged from G_1 headed in that direction. In other words, G_1 and G_2 defined a beam of molecules which went on in a vertical direction passing through G_3 into the drum if the drum was in the position shown. If the drum was not moving, the molecules struck it at P and were condensed there. If the drum was moving, then usually the

molecular beam would strike the outside of the drum, but once every revolution the slit G_3 would pass through the beam and a burst of molecules be admitted to the drum. These molecules continued to move in the straight line determined by G_1 and G_2, but by the time they got across the drum D it had turned an appreciable distance so that they struck it not at the point P opposite G_3 but at the point P' distant s from P. Suppose the molecules all had the same velocity v, then the time they used crossing the drum was $t = d/v$ where d is the diameter of the drum. If the drum is making n turns per second, then any point on its periphery is moving with a velocity πdn. Therefore,

$$s = \pi dnt = \frac{\pi d^2 n}{v} \tag{1}$$

and the velocity v of the molecules is

$$v = \frac{\pi d^2 n}{s}. \tag{2}$$

The procedure, then, would be to let molecules accumulate at P with the drum stationary long enough to make a visible deposit and then run the apparatus for a while with the drum rotating a constant number, n, of revolutions per second until a second visible deposit had accumulated at P'. Then a measurement of the distance s between the two deposits substituted in the formula for v with the known values of π, d, and n would give the velocity.

26. The procedure actually used differed only a little from this. In the first place, a segment of curved glass plate was clamped inside the drum to receive the deposit of molecules. This could be removed for observation and measurement. In the second place, the molecules do not all have the same velocity so that the deposit when the drum was rotating was not a sharp line like that at P but was smeared out into a broad band of varying thickness. The thickness of the deposit at any point was proportional to the number of molecules having the corresponding velocity so that the distance of the thickest part of the deposit from P' gives the value of the most probable velocity. The metal used by Zartman was bismuth, and some complication was introduced into his results by the fact that the molecular beam contained both monatomic (Bi) and diatomic (Bi_2) molecules.

27. The curve in Fig. 9 shows the relative thickness of the deposit plotted against displacement from P' for a temperature of the furnace E of 851°C. and a speed of rotation of the drum of 241.4 r.p.s. The corresponding molecular velocities calculated from the formula are given on the lower horizontal scale. The plus marks on the diagram are the observed thicknesses of the deposit, whereas the curve itself is drawn from the predictions of kinetic theory. (See Chap. VIII where the general question

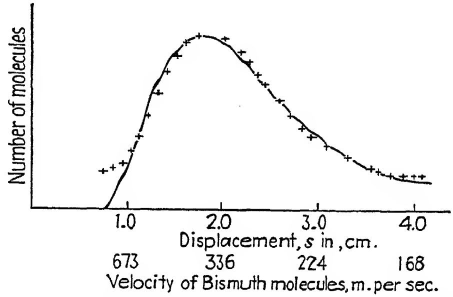


FIG. 9.—Results of Zartman's experiments on molecular velocities. The plus marks are experimental values and the curve is calculated from theory.

of the distribution of velocities among molecules is discussed.) For the present, attention need be drawn only to the value of the most probable velocity, about 340 m./sec. or about 13 mi./min., which confirms our earlier suggestion that the molecules of a gas were in rapid motion. A complete description of Zartman's apparatus will be found in the *Physical Review*, vol. 37, p. 383, 1931.

Intermolecular Forces

28. In the early part of this chapter, it was pointed out that there were two main factors controlling the behavior of the molecules of a substance, their motion and the forces between them. Phenomena have now been described that establish beyond a reasonable doubt the existence of this molecular motion and its general nature. We shall turn to the question of the forces of attraction between molecules.

29. Every known substance can be vaporized, *i.e.*, turned into a gas, if its temperature is sufficiently raised; and conversely,

every known substance can be solidified if its temperature is sufficiently lowered. The assumption of the kinetic theory is that these effects of temperature result entirely from changes in molecular motion and that the law of force between molecules is independent of temperature. Such forces should exist then in any gas and might be detected in spite of the obscuring effect of molecular agitation. Not only has this been found to be true, but it is of the greatest practical use in the liquefaction of gases. One process of making liquid air depends on the elementary physical fact that it takes energy to separate two bodies that attract each other. If air at high pressure is allowed to expand suddenly, the average separation of the molecules increases, and if there are forces of attraction between them energy must be expended. The most readily available source of energy is the energy of the molecules themselves, their energy of motion, called technically their kinetic energy. But we have seen that this increases with temperature, and conversely the temperature may be expected to decrease as the kinetic energy of the molecules is decreased. Therefore, air should cool when it is allowed to expand. It does so. It does so with such effect that it may even cool below its boiling point and condense as a liquid. The details of this process will be given later after the notions of temperature and energy have been fully discussed. A simple example of the same effect may be mentioned here however. If the air in an inflated automobile tire is let out through the valve, it will be noticed that the valve gets cold.

30. But the molecular forces are most striking when they overbalance the disruptive effect of molecular agitation, that is, in solids and liquids. They are the forces that give cohesiveness to stone and steel. That these intermolecular forces vary rapidly with distance may be seen from the breaking of any solid. Once the two parts are separated, it is usually impossible to bring the pieces together close enough for the forces to take hold again. Nevertheless, by taking very special precautions, it has been possible to heal small cracks in glass by pressing the surfaces together again. Or if two very well ground and polished flat metal or glass surfaces are squeezed together, they adhere firmly without any cement. This adhesion is caused by the intermolecular forces.

Elasticity

31. The forces between the molecules in a solid may be tested by less violent means than fracture. A steel wire may be stretched without breaking and will even return to its original shape if it has not been stretched too far. This capacity of a solid for returning to its original form after any kind of small distortion, stretching, twisting, or compression is called elasticity. It is a property possessed by all solids to a greater or less extent. Quantitatively it is found that the amount of the distortion is proportional to the force producing it. Or, in technical language, *strain is proportional to stress.* This is known as Hooke's Law. The "strain" is merely a measure of the deformation or twist produced, and the "stress" is a measure of the forces called into play throughout the whole extent of the deformed body tending to bring it back into its original shape and size, in other words, the intermolecular forces. The stress is not itself ever measured directly, but it is in equilibrium with the external forces producing the distortion and they can be measured directly. It is found that, if the external load is measured in force per unit area and the strain is expressed as a fractional change in some dimensions of the deformed body, the factor of proportion between them is a constant for any given substance and kind of deformation. These factors are called moduli of elasticity.

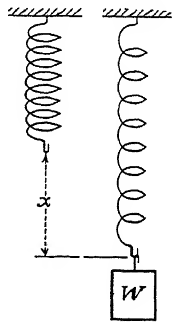


FIG. 10.—Hooke's Law. The "strain" x is proportional to the weight W.

32. It is customary in elementary textbooks to consider only one or at the most two types of elastic deformation and their corresponding moduli. There are many types of deformation but they are all manifestations of the same fundamental elastic properties of the material. We shall consider the three most commonly encountered.

1. The first type of elastic distortion is a simple compression. It can be most clearly described in terms of a uniform sphere of some solid submerged in a liquid (Fig. 11). The force on the solid is then a uniform pressure over its surface and is everywhere directed toward the center of the sphere. The surface of the sphere will be pushed inward until the elastic forces, *i.e.*, the

repulsions between adjacent molecules, combine to give forces outward along the radii of the sphere which are just equal to the inward push of the liquid. As the sphere is submerged deeper and deeper in the liquid, the hydrostatic pressure increases and the compression increases correspondingly. The amount of this compression can be measured in several ways, by the shortening of the radius of the sphere for example, but the most satisfactory way is by the change in the volume of the sphere. It is found that, *within limits, this change is proportional to the increase in pressure*, an example of the general law cited above that strain is proportional to stress. If v is the original volume and Δv is the change in volume caused by a change in pressure Δp, we can express the result as follows:

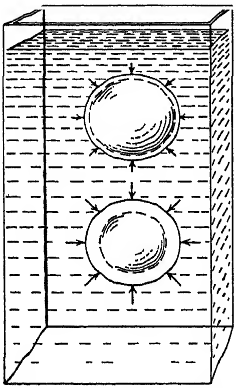


FIG. 11.—Schematic drawing of change of volume with change of pressure. The solid circles represent the size of the sphere when under no pressure; the dotted circles, its size under hydrostatic pressure at two different depths.

$$\text{Strain} = \frac{\text{change in volume}}{\text{original volume}} = \frac{\Delta v}{v}$$

$$\text{Stress} = \text{change in pressure } \Delta p.$$

Therefore

$$\frac{\Delta v}{v} \propto -\Delta p$$

or

$$\Delta p = -k\frac{\Delta v}{v} \tag{3}$$

where k is a constant of proportion known as the *bulk modulus* of the material under pressure and the negative sign shows that an increase in pressure causes a decrease in volume. The reciprocal of the bulk modulus is sometimes called the compressibility.

2. The second kind of distortion is a strain which is a deformation without a change of volume, a twisting or shearing strain as illustrated in Fig. 12. Perhaps this may be best understood by thinking of the body as made up of horizontal layers of atoms each one of which must slip a little way across the one below it to produce the deformation shown—it is this slipping of adjacent layers across each other that gives the name "shear" to this kind of strain. A certain amount of force is required to move each atom of the layer, and therefore the force required to move

the whole layer will depend on the number of atoms in it, *i.e.*, on the area of the layer which we may call A. The force on the topmost layer may be thought of as applied to the edge of a thin sheet of paper glued to the upper surface of the body. Just as the second layer of atoms exerts a force on the first layer resisting its displacement, so there is an equal and opposite force of reaction on the second layer tending to displace it. This is transmitted on down to the third layer and so on down to the bottom, which is glued to a rigid base. Each atom and each layer of atoms slips until the force from above pulling it to the right is just equal to the force below pulling it to the left. If the body is homogeneous, each layer will slip the same amount. It is found that this amount is proportional to the force applied, Hooke's law again. We must admit that our earlier statement about the nature of intermolecular forces is inadequate to explain this. It is remarkable that so simple a law should hold at all when the forces between atoms vary so rapidly with distance. It is not surprising that this simple law no longer holds if the distortion becomes at all large. To express the law of elastic shear in a form analogous to that used for compression, we should speak of changes in shape corresponding to changes in the applied force, but for simplicity in the diagram we have assumed the initial applied force as zero. Thus in the diagram, ΔX represents the distortion caused by the applied force ΔF. Δx represents the displacement of each layer of atoms relative to the one below it. Δx is evidently proportional to ΔF and also, as we have mentioned, inversely proportional to the area A. The total displacement ΔX of the top layer from its original position will depend not only on Δx but on the number of layers, *i.e.*, on the thickness T. These same statements in mathematical language are

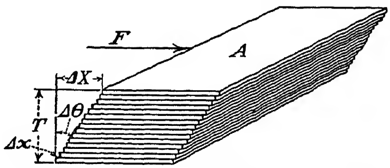


FIG. 12.—Shear diagram. (For full explanation see text.)

$$\Delta x \propto \frac{\Delta F}{A}$$

$$\Delta X \propto T\Delta x \propto \frac{T\Delta F}{}$$

Therefore,

$$\Delta F = \frac{nA\Delta X}{} \tag{4}$$

where n is a constant of proportion called the rigidity modulus. For small distortions, $\Delta X/T = \Delta\theta$ so that we can write

$$\Delta F = nA\Delta\theta. \tag{5}$$

3. The third example of deformation is probably the most familiar one. It is the change in length of a wire or beam caused by tension or compression. Here again, the process can be interpreted in terms of the motion of atoms, but this will be left to the student. The result will be expressed immediately by the equation

$$\Delta F = yA\frac{\Delta L}{L} \tag{6}$$

where ΔF is the applied force, A the cross section of the wire, ΔL the change in length of the wire, L the original length, and y the modulus, known in this case as Young's modulus.

TABLE 1.—ELASTIC MODULI
Average values of elastic constants in newtons* per square meter

Material M	Bulk modulus k	Modulus of rigidity n	Young's modulus y
Aluminum	7.6×10^{10}	2.4×10^{10}	6.0×10^{10}
Copper	14.3×10^{10}	4.2×10^{10}	12.0×10^{10}
Steel	18.1×10^{10}	8.0×10^{10}	20.0×10^{10}
Lead	5.0×10^{10}	0.54×10^{10}	1.6×10^{10}
India rubber	....	0.00016×10^{10}	0.05×10^{10}
Crown glass	5.0×10^{10}	2.9×10^{10}	7.1×10^{10}
Oak	....	....	1.3×10^{10}

* For the definition of the newton, see Chap. IV, p. 74. It is equal to 10^5 dynes.

33. From the preceding table, it is evident that large forces are required to produce any considerable amount of elastic deformation. However, if the deformation is very large, there is a permanent slipping of the body so that it does not return to its original condition after the deforming force is removed. In such cases, Hooke's law no longer applies and the elastic moduli have no meaning. In terms of atomic structure, this permanent change of shape might be expected when, in case 2, the slip Δx between adjacent layers of atoms becomes greater than the dis-

tance between adjacent atoms in a layer, for in such circumstances each atom in the upper layer might find a new position of equilibrium under the attractive forces of atoms in the lower layer farther to the right than those which originally held it. Similar explanations can be applied to the other two cases.

34. If the elastic behavior of bodies is really explicable in terms of atomic forces, there should be relations between the different moduli themselves and also between these moduli and other physical constants. This is found to be true, but the complete quantitative interpretation of these relations is so complicated a problem that many aspects of it remain unsolved.

35. In our discussion of gases, we assumed that the molecules were spherical, which implies that the forces they exert are the same in all directions. That this is not so is shown by the formation of crystals. The structure of crystals will be discussed later, but it is of interest to point out here that in crystals the values of the elastic constants depend on the direction in the crystal in which they are measured.

Surface Tension

36. In the early part of the chapter, it was suggested that the intermolecular forces in liquids were insufficient to hold the molecules in fixed positions but were strong enough to keep them all close together and to have great influence on the behavior of the liquid. This influence is most evident at the surface of liquids and at the lines of contact of these surfaces with solids. Before describing the observed phenomena, let us consider the forces on a molecule in a liquid. Suppose at first that the molecule is down in the body of the liquid completely surrounded by other molecules. It will be attracted by all the molecules in its immediate neighborhood but we have seen that these intermolecular forces fall off very rapidly with distance so that the effect of molecules more than a small distance away may be neglected. Imagine a sphere with this small distance as radius described around the molecule under consideration. Then this sphere contains all the molecules whose effect needs to be considered. Furthermore, their resultant effect is zero since they will act equally in all directions. But now suppose the molecule under consideration rises to the surface so that the sphere is half out of the water and half in. The upper half of the sphere now

contains only a few air molecules whose effect will be trifling compared with that of the densely packed water molecules in the lower half of the sphere. Consequently, there will be a resultant downward force on the molecule in the surface, due to the uncompensated attraction of the molecules immediately under it. There will be a similar force on every other molecule in the surface so that the whole surface layer is continually drawn in toward the body of the liquid. The results of this attraction are seen in the tendency of small bodies of liquid to form spherical drops. For example, mercury on a clean surface of glass draws itself together into droplets as nearly spherical as the forces of gravity will allow. Water on an oily surface will do the same thing. These effects can be described if the liquid surface is thought of as being held taut and acting somewhat as though it were a stretched elastic membrane. This point of view gives the name "surface tension" to the phenomenon.

Some Examples of Surface Tension

37. Perhaps the most familiar example of surface tension is furnished by the various kinds of small water bugs that can be seen moving about on the surface of the backwaters of any small stream. They skim about on top of the water with the greatest of ease without apparently even getting their feet wet. Their feet make little depressions in the surface but do not break through. That their feat is not one of supernatural skill is obvious when we find that an entirely inanimate object like a needle can do practically the same thing. If it is clean and dry, it can be made to float on the surface of water, though its weight makes a depression. If one end is pushed under water, then the needle sinks endwise as if it were slipping through a hole punctured in a sheet of paper.

38. Another example of the skinlike effect of a water surface can be shown with an ordinary tumbler full of water. If it is first filled to the brim, the water level can be raised until it is considerably above the edge of the glass by slipping coins or nails through the middle of the surface with as little disturbance of the surface as possible. Just before the water begins to spill, the surface can be seen to be convex as if it were a skin fastened to the edge of the glass and pushed up in the middle by the pressure of the water.

39. Perhaps the most striking simple experiment on surface tension is to make a sieve float or hold water. To do this, the sieve has to be oiled so that the water does not wet it and has to be of fairly fine mesh. Also it has to be carefully handled. To make it float, it must be lowered in gingerly fashion onto the surface. To fill it with water, a small piece of paper must be put in it to take the first shock of the water as it is poured in. In each case, the water again acts as if it had a skin which has to be stretched or jolted considerably before it will break.

40. Some of the best examples of surface-tension effects are obtained with soap films. One of these, we shall describe in detail since it gives a good idea of how surface tension may be described quantitatively. Suppose we take a rectangular wire frame such as *ABCD* in Fig. 13. Put a light cross wire *EF* on it which can slide up and down and which divides the rectangle into two parts. Dip the whole arrangement in a soap solution such as is used for blowing soap bubbles. It will come out with a continuous film filling the whole space *ABCD*. The slider *EF* can be moved back and forth as before making the spaces *AEFD* and *EBCF* larger or smaller, and it will stay in the position where it is left. Now puncture the film in *AEFD*. The slider is immediately drawn up to *BC* and if it is pulled down toward *AD* it will slide back to *BC*. This will continue until the film in *EBCF* breaks. In other words, this film is continually pulling on the slider, and unless the pull is just counteracted by that or the other film in *AEFD*, the slider will move. For the pull of the second film, we can substitute the pull of a small weight attached by a thread to the middle of *EF* and in this way can measure the pull of the film. We find that the amount of this pull depends only on the composition of the film and the length of *EF*. Now this film is of finite thickness, in fact of a thickness many times the diameter of a molecule. Therefore we are dealing with the pull not of one surface but of two, the front of the film and its back. Therefore if the length of the slider *EF* is 5 cm., we are measuring the pull of 10 cm. edge of soap film. The pull of 1 cm. length of edge is defined as the surface tension of any given

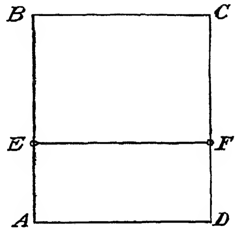


FIG. 13.—A wire frame to carry soap film for the study of surface tension.

liquid. For water at 0°C., the surface tension is 75.6×10^{-3} newton/m. This is a very small force about equivalent to the weight of a house fly.

TABLE 2.—SURFACE TENSIONS OF LIQUIDS IN CONTACT WITH AIR AT SPECIFIED TEMPERATURES
1 newton/m. = 10^3 dynes/cm.

Substance	Newtons/m.	°C.
Aluminum (molten).	0.520	750
Benzene...........	0.0289	20
Carbon tetrachloride	0.0268	20
Ether..............	0.0170	20
Ethyl alcohol.......	0.02227	20
Hydrogen..........	0.00191	−252.7
Lead (molten)......	0.422	350
Mercury...........	0.476	20
Soap solution.......	0.028	20
Water.............	0.07275	20

We can think of a water surface, then, as stretched membrane something like a drumhead. If we cut the surface anywhere, we would have to exert a force of 75.6×10^{-3} newton for every meter length of the cut to bring the two edges together again.

41. Since our interest in surface tension at this point is primarily as an example of intermolecular forces, we shall not go into a description of the various methods for getting quantitative measurements of the surface tensions of various liquids. But we shall describe some phenomena that give some indication of the relative strengths of the intermolecular forces. It is well known that glass is wetted by water but not by mercury; also, water in a small glass tube tends to run up the sides, making the surface of the water concave, whereas the edges of mercury in a similar tube will turn down, making the surface convex. This is interpreted as meaning that the forces of attraction between the glass molecules and the water molecules are stronger than the mutual attraction of the water molecules themselves, whereas the mercury molecules are more strongly pulled toward their fellows than toward the glass. A greasy surface, on the other hand, is not wetted by water so that there is no tendency for the surface skin of a body of water to be broken by contact with such a surface. Hence, we have the explanation of the role of oiling in the

sieve experiment described above. Apparently the forces between oil molecules and water molecules are less than those between glass and water and, in fact, less than those between water and water.

42. One other phenomenon should perhaps be discussed since it has been used in the determination of the size of molecules, namely, the spreading out of an oil film on a water surface. This indicates that the force between an oil molecule and a water molecule is much stronger than that between two oil molecules but weaker than between two water molecules. Thus the oil spreads out until all its molecules are as close to the water surface as possible, but they are not able in any way to disrupt this surface.

General Conclusion

43. It is hoped that the foregoing sections on the behavior of gases, solids, and liquids have shown that the atomic theory offers at least a reasonable qualitative explanation of a large group of natural phenomena. Other phenomena such as the change from liquid to solid or gas with changing temperature require more knowledge of heat and temperature than can be fairly assumed before even a qualitative explanation of them can be given. Their consideration, like the quantitative explanation of the behavior of gases which is the greatest triumph of the kinetic theory, will be postponed until after we have reviewed the fundamental principles of mechanics and discussed the general problems of temperature and heat.

SUMMARY

The kinetic atomic theory assumes that atoms exist, that they are in motion, and that there are forces between them. Their motion is shown by the Brownian movement and the diffusion of gases. Experiments on hydrogen, air, and carbon dioxide show that heavy molecules move less rapidly than light molecules. Diffusion in liquids and solids also occurs.

Molecules are very small, about 10^{-10} m. in diameter. A gas at ordinary temperature and pressure is largely empty space. Nevertheless, a molecule goes on the average only about 10^{-7} m. before colliding with another molecule in a gas at NTP. A simple experiment shows that this distance increases as the pressure decreases.

Zartman's experiment with a molecular beam and a rotating drum measures the velocity of molecules directly. These velocities are of the order of several hundred meters per second.

The cohesiveness and elastic properties of solids are examples of the effects of intermolecular forces. For small distortions, the elastic deformation of a body is proportional to the applied force. This is Hooke's law. The constants of proportion in this law are called moduli and are defined for compression, shear, and stretch.

The phenomena of surface tension are further examples of intermolecular forces.

PROBLEMS

1. If it is assumed that the force of attraction between the molecules varies as the inverse cube of the distance between molecular centers, compute the ratio of the force of attraction at 10^{-8} cm. to that at 10^{-3} mm.

2. Is there any evidence for the motion of the molecules making up a solid?

3. If a drop of oil of a volume 2×10^{-9} cu. m. covers an area of 0.2 sq. m., what is the maximum possible diameter of an oil molecule?

4. How many cubic meters are occupied by a kilogram molecule of any gas?

5. How many molecules of a gas are there in 1 cu. m. under standard conditions?

6. What fraction of the space occupied by hydrogen gas at normal temperature and pressure is taken up by the actual volume of the molecule if the diameter of the hydrogen molecule is 2.72×10^{-8} cm.?

7. The mean free path in hydrogen under normal temperature and pressure is 1.125×10^{-5} cm. What will be the mean free path in a vacuum obtained by a mercury diffusion pump at the same temperature?

8. The rate of diffusion of gases by volume is inversely proportional to the square root of their molecular weights. If it takes 5 min. for 2 l. of argon to diffuse through a small opening into an evacuated space, how long would it take 5 l. of helium to diffuse through the same opening under the same conditions of pressure and temperature?

9. What is the average distance between the centers of the molecules in a crystal of iron whose density is 7.9 g./cc.?

10. At the top of the stratosphere balloon flight of 1935 (72,395 ft.), the density of the air was one-twenty-sixth of that at sea level. What was the average distance between the molecules?

11. In an experiment with an apparatus like Zartman's, the distance PP' (see Fig. 8) was found to be 3 cm. when the drum of diameter 20 cm. was rotating at a speed of 200 r.p.s. To what molecular velocity does this correspond?

12. In Zartman's determination of molecular velocities, the drum D, 9.41 cm. in diameter, rotated with a speed of 240 r.p.s. If the drum is

rotated first in one direction and then in the other, compute the speed of the Bi molecules for which the distance between maximum deflections was 4 cm. How does this compare with the most probable velocity?

13. If the molecules of hydrogen move with an average velocity of 1,694 m./sec., what is the total distance traveled by the molecules in 1 cc. in 1 sec.?

14. There are 2.037×10^{29} collisions per second in 1 cc. of hydrogen. From the result of Prob. 13, and the fact that each collision terminates two free paths, calculate the mean free path in hydrogen.

15. If it is assumed that the attractive forces between molecules are inversely proportional to the cube of the distance between molecular centers and that the force of repulsion is proportional to the inverse fifth power of the distance, at what distance will these forces be the same? Assume ratio of the constants of proportionality of attraction to repulsion to be 10^6 when the distance is measured in centimeters.

16. Supposing the elastic laws still valid, how much pressure would be required to reduce a piece of steel to half its volume? What relation does this bear to the highest available pressure of 50,000 atm.?

17. If the compressibility of water is 2.6×10^{-10} sq. m./newton and both the compressibility of steel and water remain constant, how deep would a steel ship have to sink before it would cease sinking further?

18. A steel sphere of radius 20 cm. sinks to a depth of 1 km. in the ocean corresponding to an increase in pressure of approximately $10^3 \times 10^3 \times 9.8$ newtons/sq. m. What change is there in its radius?

19. A square block of India rubber 20 cm. on a side and 15 cm. thick has its top and bottom faces cemented to sheets of steel. How much force is required to displace the top sheet 1 cm. parallel to the bottom sheet?

20. A circular steel post used to moor a ship is 30 cm. in diameter and has a cable around it 50 cm. from the base on which the ship pulls horizontally with a force of 100 tons. Neglect the bending, and find the distance moved by the post at the point where the cable is attached under the shearing stress.

21. A steel beam of 30 sq. cm. cross section extends 60 cm. beyond the wall which supports it. If it supports a weight of 20,000 kg. on the end, what is its deflection due to shear?

22. If the steel cable used to moor the ship of Prob. 20 is 2 cm. in diameter and 100 m. long, how much will it stretch?

23. What is the area of cross section of a rubber band large enough to support the weight of a man without more than doubling in length?

24. At this latitude, the weight of 1 kg. is a force of 9.80 newtons. How far will an aluminum wire 1 mm. in diameter and 2 m. long stretch if a kilogram weight is hung on its lower end?

25. If a steel rod with a cross-sectional area of 6.45×10^{-4} sq. m. is supporting 1,500 kg., the elongation of the rod is 0.02 per cent. Find Young's modulus for the rod.

26. For a certain wire 1 m. long and 5×10^{-6} sq. m. in cross section, Young's modulus is 2×10^{11} newtons/sq. m. Find how far a mass of 10 kg. hung on the end of this wire will stretch it.

27. If a rod of elastic material is elongated 2 per cent without exceeding the elastic limit by a stress of 10^9 newtons/sq. m., what is Young's modulus?

28. A load of 10 kg. is supported by a round steel wire 1 mm. in diameter. Find the stress in the wire.

29. If water rises 1.544 cm. in a capillary tube 1 mm. in diameter, what is the surface tension of the water?

30. A pencil 16 cm. long floats on water whose surface tension is 0.073 newton/m. If a layer of alcohol is poured on the water on one side of the pencil, thus reducing the average surface tension on that side to 0.038 newton/m., find the resultant force on the pencil.

31. A match stick 5 cm. long floats on water. Ether is poured on the water on one side of the stick. What force tends to move the stick and in which direction?

32. If each foot of a six-legged water bug is round and of area 0.4 sq. mm., how much may the insect weigh before it is unsafe for him to venture on a water surface?

33. If the slider *EF* in Fig. 13 is 3 cm. long, what is the force on it when *AEFD* is empty and *BEFC* is filled with a soap film?

34. How much weight per centimeter of a slider can be supported on a soap film in a vertical wire frame such as the one in Fig. 13?

CHAPTER III

VECTORS AND LINEAR MOTION

1. In the preceding chapters, we have spoken of the forces between atoms and of the masses of atoms without any precise statement of the meaning of the terms "force" and "mass." It was assumed that the student was more or less familiar with these terms from ordinary experience and from previous study. But much of the beauty and power of physics lies in its precision and exactness. The physicist is not satisfied with a vague qualitative description and explanation of a natural phenomenon. The phenomenon must be described in terms of measurement; the characteristics of an event must be represented by numbers. Similarly, the explanation of a phenomenon must be expressible in precise language, usually the language of mathematics, before it is considered completely satisfactory. It is often impossible to do this, and it is more often impossible to do it in a form intelligible to the beginning student. But the ideal remains clear, and until the student has achieved some familiarity with rigorous quantitative reasoning he knows nothing of the methods of physics. Reasoning cannot be rigorous and quantitative if the terms used are not precisely defined. Consequently we must try to give exact definitions of terms as soon as the concepts they represent are reasonably clear. The concepts of force and of mass are of fundamental importance in every field of physics; so are the concepts of position, motion, and rate of change of motion. It is with these ideas that this chapter and the next are concerned. It will develop them in connection with a discussion of mechanics, the oldest field of physics and the one in which precision and rigor can be achieved most easily by the elementary student.

Force and Mass

2. The notion of force arises directly out of the sensation of muscular effort. We know very well what is a push or a pull and can even form a fair idea of the relative magnitudes of

various pushes and pulls. It takes a bigger force to push an automobile than a baby carriage. It takes more of a pull to get a sled started if it has two people on it than if it has only one. To serve a fast ball in tennis, one has to hit with greater force than to serve an easy ball. It takes more force to lift two suitcases than one. It is harder to stretch a thick elastic band than a thin one. Can any of these examples be used to establish a precise quantitative standard of force? All but the last one involve motion and inertia neither of which has been discussed. An elastic or spring might serve admirably were it possible to construct a standard one whose properties would not change with time; but this is not practicable. Nor is our direct muscular sensation a suitable standard of measurement. It is too subjective and too dependent on varying conditions. It would seem therefore that a qualitative intuitive notion of force is easy to get but that a precise definition must wait until standards of inertia and motion have been set up.

3. In the case of mass, the situation is almost exactly reversed. The mass of a body is usually defined as the quantity of matter in the body, but we have no way of ascertaining what this means by direct sense perception. If we could see the atoms in the body, we might count them and in that way get a number expressing the quantity of matter in the body, but we would still have to specify whether the atoms were of hydrogen or lead or some other element. We can push or pull the body or lift it. The more it resists being moved or the heavier it is, the greater we say its mass is, but these pushings, pullings, and liftings give direct sensations of force, not of mass. The more massive a body is, the greater is its inertia, *i.e.*, the greater the force required to set it in motion, and the greater is its weight, *i.e.*, the force required to lift it. In other words, masses can be compared or measured only by using forces. But it is comparatively easy to set up a standard of mass. Experience shows that the mass of a nonvolatile nonoxidizing body does not change appreciably over long periods of time so that almost any body of such a type can be taken as a standard of mass. We are confronted with a choice. Either we can set up an arbitrary standard of force and define a standard of mass in relation to the standard force, or we can set up a standard of mass and define a standard force in relation to the standard mass. The first alternative has an advantage

because of its direct relation to perception, but we have seen that it is practically unsatisfactory. Therefore we choose the second alternative. We decide to call a particular chunk of matter the standard mass, to measure all other masses by comparing them with the standard, and to measure forces by their effect on known masses.

4. The particular chunk of matter chosen to be the international standard of mass is a block of platinum-iridium alloy known as the International Prototype Kilogram and kept at the International Bureau of Weights and Measures at Sèvres, near Paris. This mass was originally intended to be exactly the equivalent of the mass of one one-thousandth of a cubic meter or 1,000 cc. of pure water at a temperature of 4°C. Because of the experimental difficulties of establishing this equivalence exactly, the block of metal is itself taken as the standard but the equivalence is sufficiently close to be of great practical convenience.

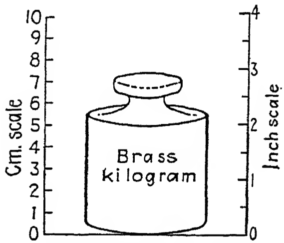


FIG. 14.—Relative sizes of a cm. scale, an inch scale, and a brass kg. mass. Reduced in reproduction to exactly ⅓ actual size.

5. In the meter-kilogram-second system used in the present book, the standard kilogram is itself taken as the fundamental unit of mass. In the c.g.s. system of units, the fundamental unit of mass is the gram, which is one one-thousandth of the mass of the standard kilogram. In this country, the more familiar unit of mass is the pound avoirdupois. It is legally defined as 0.4535924 kg. It is rarely used in scientific work.

6. Once we have a standard of mass established, the next problem is to consider how other masses can be compared with it. We have pointed out that there is no means of perceiving mass apart from force. There are two ways in which masses can be compared by using forces. The first is to study the way in which the application of a given force affects the motion of the masses to be compared. In other words, to study the inertial resistance the bodies offer to any change in their state of motion. This has theoretical advantages, as we shall see, but it is not very practicable. The other method is to compare the force of gravity on the bodies—in short, to weigh them. This is the method that is invariably used in practice. In fact, it is so familiar that weight

and mass are often confused. Actually they are by no means the same thing. The weight of a body is the gravitational force of attraction between the body and the earth. It depends on the mass of the body, the mass of the earth, and the distance of the body from the center of the earth. At a given point on the earth's surface, the last two factors are constant so that the weights of two bodies are proportional to their masses in the same way. But the distance from the earth's center varies from place to place on the earth's surface so that the pull of a body on a

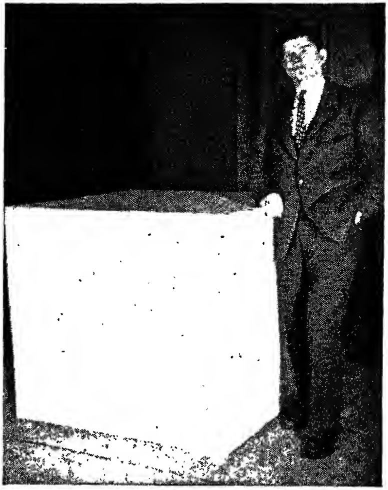


FIG. 15.—Comparison between a cube one meter on each edge and a man.

spring balance on a mountain top at the equator is appreciably less than at sea level at the pole. The weight of a body, then, is a force whose magnitude varies with position in space, but the mass of a body as determined by its inertia is found to be the same at all points in space.

7. Since the weight of a body is a force, it is possible to use the weight of a standard body as the standard of force. On this basis, the weight of a kilogram could be taken as the unit of force; or the weight of a pound could be chosen if a more traditional unit was desired. From the physicists' point of view, there are two objections to a system of this sort. In the first

place, it leads to a confusion of words. Unless one is very careful always to specify a force as so many kilograms weight or so many pounds weight, one is in the position of using the same name for units of entirely different quantities, mass and weight. In the second place, there is the objection that we have chosen a unit that is not constant but depends on the location of our standard in space. This difficulty can be overcome by choosing the weight of a given mass at a particular location as the standard of force. Nevertheless, to the physicist or astrophysicist, such a standard is unsatisfactory. The physicist therefore prefers to define a standard of force in terms of the inertia of the standard mass. To understand how this is done, we must consider at length the nature of motion, a subject of primary importance for its own sake.

Position and Displacement

8. When a body has moved, it has changed its position so that the first step in the study of motion is to learn how to specify the position of a body exactly. Such a position can be specified only with reference to some other body or bodies. For example, the position of the Naval Observatory in Washington is completely specified terrestrially by the statement that it has a latitude of 38°55′ north of the equator, a longitude of 77°4′ west of Greenwich, and an elevation of 86 m. above sea level. Again the position of a New York office might be specified by saying it was on the twentieth floor of a building at 57th Street and 5th Avenue. In each case, a position has been described by three numbers. We find that the position of a point in three-dimensional space can always be completely specified by three numbers, and we call these numbers the coordinates of the point. We call the point where all three of these numbers are zero the origin of the system of coordinates. If we are dealing with only two dimensions, only two coordinates are required. For instance, in the examples given above, we might very well have been interested only in the positions on the earth's surface and not in the height above it. Similarly in a space of four or more dimensions, four or more coordinates are required.

9. Suppose the center of the Capitol at Washington is chosen as the origin of a coordinate system. Then imagine a line drawn through the Capitol in a north-south direction and another

drawn through it in an east-west direction. These two straight lines at right angles to each other then form the axes of a two-dimensional rectangular coordinate system. Let us call the east-west line the x-axis and the north-south line the y-axis. Then we can specify the position of any point in the city, indeed of any point in the country, by saying that it is so many meters north or south of the x-axis and so many meters east or west of the y-axis. Our usual convention is to call a distance measured parallel to the x-axis, *i.e.*, perpendicular to the y-axis, an x-coordinate and one parallel to the y-axis a y-coordinate. We call

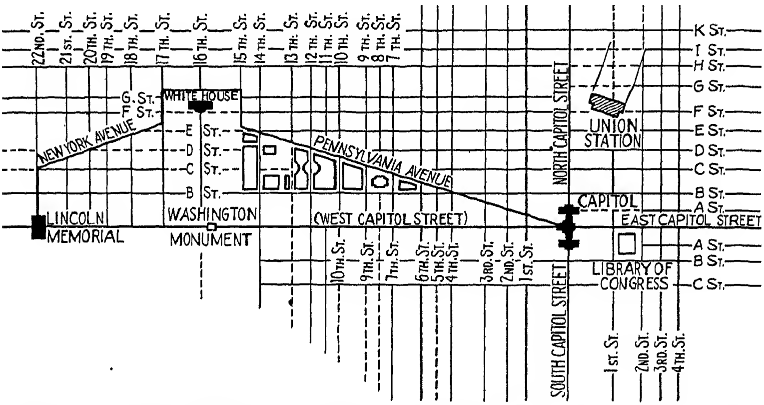


FIG. 16.—A map of the city of Washington somewhat altered to emphasize the regularity of plan.

the x positive if it is to the east and negative if to the west of the Capitol; similarly y's to the north are positive and to the south, negative. This corresponds to the general convention of plus and minus in terms of right and left and up and down in any rectangular coordinate system of this sort. We can measure the coordinates in feet, meters, miles, kilometers, blocks; or any other convenient unit of length. In fact, the system of naming streets in Washington is essentially based on the notion of a coordinate system. When we say that the White House is at 16th and G Street N.W. we mean it has coordinates $x = -16$ blocks and $y = +7$ blocks. We can say equally well that the White House is about 1.3 mi. west and 0.5 mi. north from the Capitol. When converted to meters, this gives the coordinates of

the White House as $x = -2092$ and $y = 805$ m. Another way to specify the position of the White House relative to the Capitol is by saying it is 2,240 m. in a direction 21° north of west from the Capitol. Again we are using two coordinates, but one is an angle instead of a distance.

10. Now that we have learned how to specify position exactly, we can proceed to a discussion of change of position, technically called displacement. Suppose a car starts from the Capitol and drives 5 km. in a straight line, where does it get to? No one knows, unless he also knows in what direction the car was driven. The change of position has not been adequately described. In other words, to specify a displacement completely not only the numerical magnitude (5 km. in this case), but also the direction must be given. *Quantities* of this sort *which require both a magnitude and a direction for their complete specification are called*

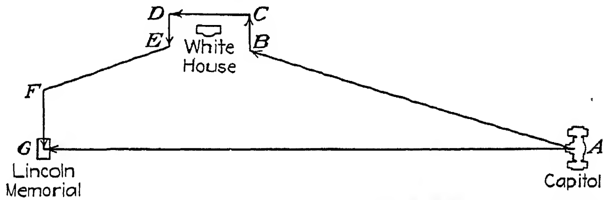


FIG. 17.—Vector diagram of displacements made in going from the Capitol to Lincoln Memorial.

vector quantities or, more briefly, vectors. They are quite common in physics (*e.g.*, force, velocity, electric field strength). By the direction of the vector, we mean not only the line along which it lies but which way it points on this line. This is called the sense of the vector and is sometimes given separately. Quantities that have no directional property, which are completely described when their magnitude is given, are called scalar quantities or scalars (*e.g.*, mass, energy, temperature).

11. Suppose that a car instead of attempting the impossible feat of driving 5 km. in a straight line from the Capitol actually carries a party of tourists from the Capitol to the Lincoln Memorial by way of the White House. Let the car stop every time it has to change its direction. Each separate stage of its journey is a movement in a certain direction for a certain distance, in short, a displacement, so that the resultant movement from the Capitol to the Lincoln Memorial can be thought of as the effect

of adding a whole set of displacements together. A schematic diagram of the route (subject to alterations by police regulations) is shown in Fig. 17. Each straight portion of the line $ABCDEFG$ is drawn in the direction of a particular part of the motion and of a length proportional to the distance traversed in that part of the motion. Thus the lines AB, BC, etc., represent the separate displacements whose added effect is the displacement AG. The arrow points on the vectors indicate that the motion is from A to B, not B to A, etc. The lines AB, BC, etc., are called vectors, and AG, itself a vector, is called the resultant of the vectors AB, BC, etc. In constructing this simple diagram, we have performed the operation known as vector addition. It can be done in a similar way for any vector quantities whether they be displacements, forces, or something else. The parallelogram of forces usually taught in school physics is a particular case where only two vectors are combined to get a single resultant.

Displacement in Terms of Coordinates

12. It is sometimes convenient to express a displacement, or at least its magnitude, in terms of the coordinates of the beginning and end points. This is done very simply by applying the familiar theorem of geometry that the square of the hypotenuse of a right triangle is equal to the sum of the squares of the two other sides. Suppose the two points P_1 and P_2 in Fig. 18 have the coordinates (x_1y_1) and (x_2y_2). Then the displacement P_2P_1 is the hypotenuse of the right triangle P_2P_1Q, and therefore

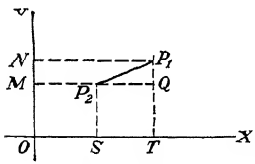


FIG. 18.—Displacement in terms of coordinates (see text).

$$\overline{P_2P_1}^2 = \overline{P_2Q}^2 + \overline{QP_1}^2$$

but $ON = y_1$ and $OM = y_2$. Therefore, $QP_1 = MN = y_1 - y_2$. Similarly, $P_2Q = x_1 - x_2$. Substituting these values in the equation for $\overline{P_2P_1}^2$ and taking the square root, we have for the distance between any two points in terms of their coordinates

$$P_2P_1 = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{\frac{1}{2}}.$$

This may be easily generalized to three dimensions giving

$$[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2]^{\frac{1}{2}}$$

for the distance between the point $(x_1y_1z_1)$ and the point $(x_2y_2z_2)$.

Vectors

13. The general procedure of vector addition is as follows: Take any vector quantities A, B, C, etc., whose resultant effect is desired. They may be displacements, velocities, forces, or something else, but they must all be quantities of the same kind. Starting from some fixed point on a sheet of graph paper, draw a line in the direction of A and of length proportional to the size of A. From the end of this line, draw another in the direction of B and of length proportional to the size of B. Keep on adding straight segments of line in the directions of the successive vectors and proportional to their sizes. It makes no difference in what order the vectors are taken. When they have all been represented, connect the starting point to the end of the last vector. This will give a line in the direction of, and of length proportional to, the resultant effect of all the separate vector quantities with which we started. An example of this process has already been given in Fig. 17. A general example is shown in Fig. 19.

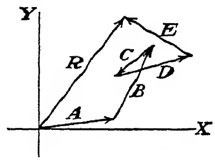


FIG. 19.—A polygon of vectors.

14. The object of vector addition is to find a single vector that has the same resultant effect as the several vectors given initially. Sometimes it is desirable to reverse the process, *i.e.*, to find several vectors that will give the same resultant effect as a single initial vector. In particular, it is often desirable to replace a vector by two vectors at right angles to each other and parallel to the coordinate axes. This process is called resolving a vector into its components. In Fig. 20, OP is a vector drawn out from the origin and making an angle θ with the x-axis. Drop a perpendicular from P to the x-axis at A. Then the vector AP parallel to the y-axis added to the vector OA parallel to the x-axis clearly gives the resultant vector OP. OA and AP are called the x- and y-components of OP. Recalling the definitions of the sine and cosine in trigonometry, we see that

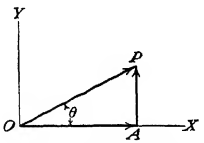


FIG. 20.—The resolution of a vector into two components.

$$OA = OP \cos\theta$$

and (1)

$$AP = OP \sin\theta.$$

Also,

$$\overline{OA}^2 + \overline{AP}^2 = \overline{OP}^2 \tag{2}$$

and

$$\frac{\overline{AP}}{\overline{OA}} = \tan\theta. \tag{3}$$

15. One great advantage of this process of resolving vectors into components along the axes is that it makes it possible to replace a system of vectors oriented in all directions by two sets of vectors, one set parallel to the x-axis and one set parallel to the y-axis. These can then be combined to get the resultant of the original system of vectors. The process is as follows:

16. Take the vectors A, B, C, etc., making angles θ_1, θ_2, θ_3, . . . with the x-axis. Then resolve A into its components A_x and

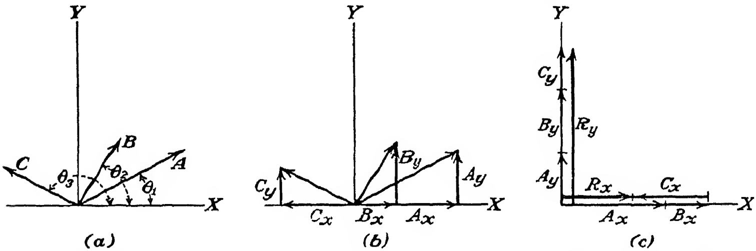


FIG. 21.—The reduction of three vectors to two components at right angles.

A_y along the x- and y-axes. Similarly, resolve the other vectors into their components B_x and B_y, C_x and C_y, etc. In this way, any number of vectors A, B, C, etc., can be replaced by an equal number of vectors along the x-axis plus an equal number of vectors along the y-axis. But all the vectors along the x-axis have the same direction, and therefore their resultant is the algebraic sum of their magnitudes so that we can write

$$R_x = A_x + B_x + C_x + \ . \ . \ .$$

and similarly (4)

$$R_y = A_y + B_y + C_y + \ . \ . \ . \ .$$

Thus the original vectors A, B, C, etc., are equivalent to two vectors R_x and R_y along the x- and y-axes. But these in turn may be combined by vector addition to one resultant vector R. Further, we know that

$$R^2 = R_x^2 + R_y^2 \tag{5}$$

and

$$\tan\theta = \frac{R_y}{R_x} \tag{6}$$

But we also know that

$$A_x = A\cos\theta_1, \qquad B_x = B\cos\theta_2, \qquad C_x = C\cos\theta_3,\ \text{etc.},$$

and (7)

$$A_y = A\sin\theta_1, \qquad B_y = B\sin\theta_2, \qquad C_y = C\sin\theta_3,\ \text{etc.}$$

Therefore,

$$R_x = A\cos\theta_1 + B\cos\theta_2 + C\cos\theta_3 \ldots$$

and (8)

$$R_y = A\sin\theta_1 + B\sin\theta_2 + C\sin\theta_3 + \ldots,\ \text{etc.}$$

17. Therefore if we know the magnitudes and directions of a number of vectors, we can calculate the magnitude and direction of a single vector, called the resultant, which has the same effect as all the original vectors combined. This is equivalent to the process that was performed by graphical means in Par. 11 for displacements. The analytical method just described is usually easier to apply and more accurate.

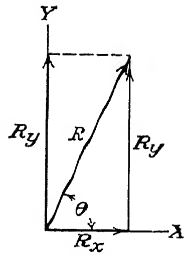


FIG. 22.—Composition of two rectangular components into a single resultant.

Velocity

18. Position and change of position have been discussed, and a digression has been made to give some account of vector quantities. Now we shall return to the development of the topic of motion by considering rate of change of position with time, known technically as velocity. In considering displacement, we pointed out that both a magnitude and a direction had to be specified. We could not give the location of a car merely by saying that it had gone 5 km. in a straight line from the Capitol. We had to specify the direction of the motion as well. In the same way, it makes a great deal of difference whether a car moving with a velocity of 50 mi./hr. is headed straight down a road or at an angle that will land it in the ditch. Furthermore, unless the road is perfectly straight, changes in the direction of

the car's velocity will frequently be necessary although the magnitude of the velocity, what we call the speed of the car, may not be changed. Velocity, like displacement, is a vector quantity involving both a magnitude and a direction.

19. The only measurements required in the study of displacements are measurements of length. A steel tape would be sufficient equipment. We have been using various units of length in our descriptions according to convenience, but, as we explained in our introductory note on units, we are taking the meter as our basic unit. It is defined as the length of the standard meter kept in the International Bureau of Weights and Measures at Sèvres, near Paris. To measure velocities, we must also be able to measure time. Fortunately there is only one system of time units in common use, and it is perfectly familiar. We shall take the ordinary second as our standard unit of time and shall leave its exact definition in terms of the motion of the heavenly bodies to the astronomers. Evidently the scientific unit of velocity is the meter per second in the m.k.s. system or the centimeter per second in the c.g.s. system. Neither of these units is familiar in ordinary experience. We have very little idea whether a car going 5 m./sec. is moving fast or slowly. It is only when we convert the velocity to $5 \times 60 \times 60 = 18{,}000$ m./hr. and finally to $18{,}000/1{,}000 = 18$ km./hr. that we realize the car is going at moderate speed. Even then, we may not feel quite sure about it until we have looked up in a table to see how many kilometers there are in a mile (1.6). For this reason, we shall use miles and hours as the units of distance and time in our first example of velocity.

20. By way of illustrating the measurement of a velocity, consider a car traveling along the straight road between Philadelphia and New York known as the Brunswick Pike. Take the traffic circle at the village of Penns Neck as the origin of coordinates, and call the road itself the x-axis. Suppose that at two o'clock a car is 5 miles along the road toward New York and that at quarter past two it is 15 miles along the road toward New York. Its average velocity during that quarter of an hour must have been 40 mi./hr. toward New York. In arriving at this conclusion, we have not had to worry about direction since the car has been confined to the straight road. The calculation of the magnitude of the velocity was as follows:

$$\frac{15 - 5}{2\frac{1}{4} - 2} = \frac{10}{\frac{1}{4}} = 40$$

and illustrates the general method for determining a speed. Thus, for a body moving parallel to the x-axis, if its position A at the time t_0 is specified by x_0 and its position B at the time t by x, its average speed in the meantime is given by

$$v_{av} = \frac{x - x_0}{t - t_0}.$$

If the motion is not parallel to an axis, the equation must be generalized to

$$v_{av} = \frac{\sqrt{(x - x_0)^2 + (y - y_0)^2}}{t - t_0}$$

for two dimensions or to

$$v_{av} = \frac{\sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}}{t - t_0}$$

for three dimensions. These relations give the magnitude of a uniform velocity which would carry the body in a straight line from A to B in the time $t - t_0$. It is thus an average as to both magnitude and direction.

21. Often it is not the average velocity over a period of time that interests us but the instantaneous velocity, and it is desirable that we should have a clear-cut definition of what this is. In the illustration given, it is likely that the speed of the car fluctuated considerably—perhaps from 20 to 60 mi./hr. as shown by the speedometer. Had we chosen a time interval of 1 min. instead of 15, the probability of fluctuations would have been less. An interval of 1 sec. would have been still better. In principle, we could choose the interval as small as we like, infinitesimally small in fact, and the distance covered in that time would be correspondingly small; but the ratio between the two would remain large and would represent the instantaneous velocity at that particular instant. Symbolically, this is written

$$v = \frac{\Delta x}{\Delta t} \tag{9}$$

where Δt and Δx are extremely small compared with the times and distances involved in the problem under consideration. Students

of the calculus will recognize that Δx and Δt may be allowed to approach zero without any loss in the significance of their ratio. Equation (9) is then written

$$v = \frac{dx}{dt}.$$

Acceleration

22. As has been pointed out in the previous paragraphs, the velocity of a moving object will not, in general, be constant. Just as the position of the body changes with time, so may the velocity of the body change with time. The rate of change of velocity with time is called the acceleration. The ideas of average acceleration and of instantaneous acceleration can be developed and defined by a treatment analogous to that used for velocity. In fact, this treatment can be easily applied to any property of a body that is changing with time, such as temperature or electrical resistance. Thus, suppose the velocity of a body at a time t_0 is v_0 in a certain direction and that at a later time t it is v in the same direction; then the average acceleration during the interval is given by

$$a_{av} = \frac{v - v_0}{t - t_0}$$

The instantaneous acceleration is obtained by allowing the time interval to become infinitesimal, giving

$$a = \frac{\Delta v}{\Delta t} = \frac{dv}{dt}. \tag{10}$$

The acceleration may itself vary but, in this book, we shall concern ourselves only with cases where the acceleration remains constant, at least in magnitude, except in the study of simple harmonic motion. This limitation is partly because the mathematical difficulties in the handling of variable accelerations are considerable and partly because the acceleration is constant in many of the most important problems of physics.

23. The notion of acceleration is familiar to any driver of a car, but he usually is not accustomed to expressing it very precisely. Thus, suppose the speedometer of a car reads 30 mi./hr. and a minute later reads 40 mi./hr. Scientifically, we would say that its average acceleration during that minute had been 10 mi./hr./min.; that is, a change in velocity of 10 mi./hr. has occurred in

1 min.; the change per second is one-sixtieth as much, and therefore the acceleration might be equally well expressed as $\frac{1}{6}$ mi./hr./sec. Similarly, the change in velocity need not be given in miles per hour. Ten miles per hour is equivalent to 14.7 ft./sec. so that the acceleration could be expressed as 14.7 ft./sec./min. Acceleration, like velocity, is a vector quantity. If the acceleration is in the same direction as the velocity, the velocity is increasing in magnitude but not changing in direction. If the acceleration is negative, *i.e.*, in the opposite direction to the velocity, the velocity is decreasing (*e.g.*, when the brakes of a car are applied). If the acceleration is at right angles to the

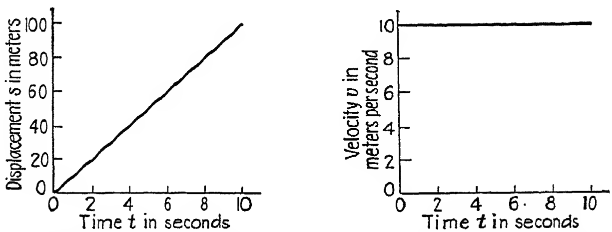


FIG. 23.—Displacement and velocity as a function of the time with no acceleration. $a = 0$, $s = vt$.

velocity, the velocity is changing in direction but not in magnitude (*e.g.*, in uniform circular motion). If the velocity is changing in both direction and magnitude, the acceleration is at an angle to it. The most familiar example of this general case is a projectile. In dealing with such problems, we shall see that the possibility of resolving vectors into components is of the greatest assistance. For the present, we shall confine ourselves to the case of motion in a straight line with constant acceleration.

Uniformly Accelerated Linear Motion

24. If, at the time $t = 0$, the distance of a moving body from some point of reference is s_0, then at any other time t its distance s from the reference point is given by

$$s = \quad + v_{\text{av}}t \tag{11}$$

where v_{av} is its average velocity during the time t. This follows directly from the definition of velocity. In most problems, the

point of reference can be chosen as the point where the body is when $t = 0$ and, therefore we have

$$s = v_{av}t. \tag{11'}$$

If the velocity is constant, this becomes $s = vt$. If the body is moving with constant acceleration a and its initial velocity at

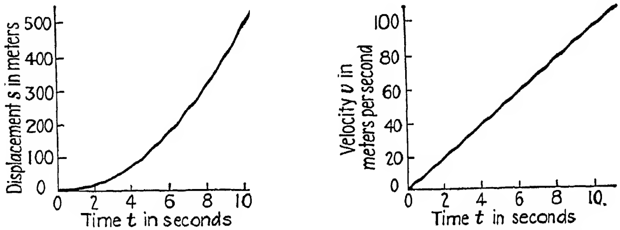


FIG. 24.—Displacement and velocity as a function of the time with constant acceleration $a = 9.8$ m./sec.[2] The initial velocity is zero. $v = at$, $s = \frac{1}{2}at^2$.

$t = 0$ is v_0, then its velocity at any other time t is

$$v_t = v_0 + at \tag{12}$$

and its average velocity in the time interval $t - t_0$ is

$$v_0 + v_t$$

and therefore

$$v_{av} = v_0 + \frac{at}{2}.$$

Substituting this in (11′), we get

$$s = v_0t + \tfrac{1}{2}at^2. \tag{13}$$

Sometimes problems arise in which we want to get direct relations between s, v_0, v_t, and a, without concerning ourselves with the time. By eliminating t between Eqs. (12) and (13), we get the equation

$$2as = v_t^2 - v_0^2. \tag{14}$$

If we group together Eqs. (12), (13), and (14), we have all the relations that we need for the study of uniformly accelerated motion. These equations are used repeatedly in many parts of

physics and should be thoroughly familiar to the student. For convenience, they are gathered together as follows:

$$v_t = v_0 + at \tag{12}$$

$$s = v_0 t + \tfrac{1}{2}at^2 \tag{13}$$

$$2as = v_t^2 - v_0^2 \tag{14}$$

(sometimes written $v_t = \sqrt{v_0^2 + 2as}$).

SUMMARY

We have direct sense perceptions of force but not of mass. Mass is experienced only indirectly through its inertial reaction against acceleration or through its weight. Nevertheless, a standard of mass is set up rather than one of force. The standard of mass is a certain block of platinum-iridium alloy which is defined as one kilogram. Masses are usually compared by weighing.

The position of any point in space is completely specified by three numbers called coordinates. A change of position is called a displacement. Displacement is a vector since vectors are defined as quantities that require both a direction and a magnitude for their complete specification. Quantities that have no directional properties are called scalars. A displacement can be expressed in terms of the coordinates of the initial and final positions of the body displaced.

The resultant of two vectors such as two displacements or two forces acting simultaneously can be found by the process of vector addition. This process can be carried out either geometrically or analytically. A vector can be resolved into two component vectors at right angles to each other whose resultant effect is equivalent to that of the original vector. If the original vector A makes an angle θ with the x-axis, the x-component is $A \cos \theta$ and the y-component is $A \sin \theta$.

The velocity of a particle is the rate of change of the position of that particle. Velocity is a vector. Acceleration is the rate of change of the velocity. Acceleration is also a vector.

From the definitions of displacement, velocity, and acceleration, the following relations among them are derived for motion along a straight line:

$$v_t = v_0 + at$$

$$s = v_0 t + \tfrac{1}{2}at^2$$

$$2as = v_t^2 - v_0^2$$

where v_0 and v_t are the velocities at the time zero and t, respectively, s is the displacement during that time, and a is the acceleration assumed constant.

ILLUSTRATIVE PROBLEMS

1. A small trading ship is pursuing an irregular course from port to port in a group of islands. It steams at the rate of 12 knots as follows: two hours due east, half an hour northwest, one hour 15 degrees south of east, and finally three hours southwest. At the end of its trip, how far is it from its starting point and in what direction?

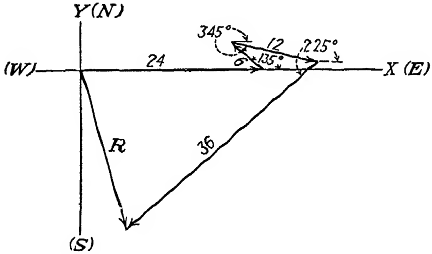


FIG. 25.—Displacements of trading ship. Vector addition

This problem clearly deals with change of position, *i.e.*, displacement, but the data needed to solve the problem are given in terms of speed and time. The first thing to do is to reduce the data to a more useful form by multiplying the times by the speed and referring all the directions to the same standard. If we call the east-west line the x-axis and the north-south line the y-axis, the displacements are

Nautical Miles	
24.........	Angle of 0° with the positive direction of the x-axis
6.........	135°
12.........	345°
36.........	225°

and are shown graphically in Fig. 25. The problem is to find the resultant of all these displacements, in other words, to add them vectorially. To do this, we first find all the x-components and add them algebraically and then do the same for the y-components. Thus:

x-components

$24 \cos 0° = 24 \times 1$		24
$6 \cos 135° = 6 \times (-0.707)$		-4.24
$12 \cos 345° = 12 \times 0.966$		11.6
$36 \cos 225° = 36 \times (-0.707)$		-25.45
	$R_x =$	$+5.91$

$$
\begin{array}{rcll}
 & & y\text{-components} & \\
24 \sin 0^\circ & = & 24 \times 0 & = \quad 0 \\
6 \sin 135^\circ & = & 6 \times 0.707 & = \quad 4.24 \\
12 \sin 345^\circ & = & 12 \times (-0.259) & = \; -\ 3.10 \\
36 \sin 225^\circ & = & 36 \times (-0.707) & = \; \underline{-25.45} \\
 & & R_y & = \; -24.31
\end{array}
$$

By this process, we have found two displacements R_x and R_y at right angles to each other which are equivalent to the original four displacements at various angles. The next step is to combine these two to find the resultant displacement. The magnitude of the resultant is

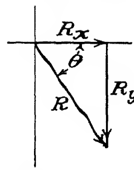


FIG. 26.—Addition of displacements at right angles to each other.

$$R = \sqrt{R_x^2 + R_y^2} = \sqrt{625.904} = 25.02.$$

Since R_x is positive and R_y is negative, the resultant is in the fourth quadrant and the angle it makes with the x-axis is

$$\theta = \arctan \frac{R_y}{R_x} = \arctan 4.113 = 76°20'.$$

Therefore, the steamer at the end of its trip is 25.02 nautical miles in a direction 76°20′ south of east from its starting point.

2. Suppose that the steamer of Prob. 1 on the last leg of its trip got into an ocean current of three knots flowing toward the east. In what direction must it have headed in order to continue to move toward the southwest, and how fast did it move in that direction?

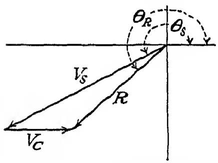


FIG. 27.—Vector addition of velocities.

In this problem, the direction of the resultant velocity, the magnitude and direction of one component (the ocean current), and the magnitude of the other component (the vessel's speed) are given. The direction of the second component and the magnitude of the resultant are to be determined. First draw a diagram to get the situation clearly in mind. Thus, we are given θ_R, V_c, θ_c $(= 0)$ and V_s, and we have to find θ_s and R. Now follow the usual procedure, and replace these vectors by an equivalent set of vectors parallel to the x- and y-axes. The x-components are

$$
\begin{aligned}
R_x = R \cos \theta_R &= R \cos 225^\circ = -0.707R \\
V_c \cos \theta_c &= V_c \cos 0^\circ = 3.0 \\
V_s \cos \theta_s &= 12 \cos \theta_s
\end{aligned}
$$

and we know that

$$R_x = (V_c)_x + (V_s)_x \qquad \text{or} \qquad R_x = 3 + 12 \cos \theta_s.$$

Similarly,

$$R_y = V_s \sin \theta_s = 12 \sin \theta_s$$

But since $\theta_R = 225°$ and $\tan 225° = \tan 45° = 1$, we know that

$$\frac{R_y}{R_x} = 1 \qquad \text{or} \qquad R_x = R_y$$

therefore

$$12 \sin \theta_s = 3 + 12 \cos \theta_s$$

an equation that we have to solve for θ_s. Rewriting, we have

$$12(\sin \theta_s - \cos \theta_s) = 3.$$

Squaring gives

$$144(\sin^2 \theta_s - 2 \sin \theta_s \cos \theta_s + \cos^2 \theta_s) = 9$$

but $\sin^2 \theta + \cos^2 \theta = 1$ and $2 \sin \theta \cos \theta \quad \sin 2\theta$.

Therefore

$$144(1 - \sin 2\theta_s) = 9$$

or

$$-\sin 2\theta_s = \frac{9}{144} - 1 = -\frac{135}{144}$$

$$2\theta_s = \arcsin \frac{135}{144} = \arcsin 0.9375$$

$$2\theta_s = 69°38' \text{ or } 69°38' + 360°.$$

Therefore

$$\theta_s = 34°49' \text{ or } 214°49'$$

and the second value must be correct since R is in the third quadrant. Furthermore,

$$R \quad \frac{12 \sin \theta_s}{0.707} \quad \frac{12 \times 0.5709}{0.707} = 9.69.$$

Therefore the ship must head 34°49′ south of west and will then move in a southwesterly direction at a rate of 9.69 knots.

3. A train going at 60 mi./hr. applies the brakes ½ mi. from the station. (*a*) What is its deceleration? (*b*) After leaving the station, the train accelerates at the rate of 15 mi./hr./min. How far will it travel before reaching the speed of 60 mi./hr. again? (*c*) Allowing 2 min. for the stop at the station, how much time elapses between the moment when the brakes are applied and the time when the train is going at full speed again after the stop?

a. Distance, initial velocity, and final velocity are given and acceleration (negative) is asked for; therefore apply Eq. (14). We have

$$v_0 = 60 \text{ mi./hr.}$$
$$v_t = 0$$
$$s = 0.5 \text{ mi.}$$

so that

$$\overline{2s} \quad \frac{-(60 \text{ mi./hr.})^2}{1 \text{ mi.}} \quad -3{,}600 \text{ mi./hr.}^2$$

$$-60 \text{ mi./hr./min.},$$

i.e., the train decelerates at the rate of 60 mi./hr./min.

b. Here we are given the initial velocity, the final velocity, and the acceleration and are asked for the distance. Again we apply Eq. (14), but now

$$\begin{aligned} v_0 &= 0 \\ v_t &= 60 \text{ mi./hr.} \\ a &= 15 \text{ mi./hr./min.} = 900 \text{ mi./hr.}^2 \end{aligned}$$

Hence,

$$\frac{(60 \text{ mi./hr.})^2}{2 \times 900 \text{ mi./hr.}^2} = 2 \text{ mi.}$$

c. The data are the same as for (*a*) and (*b*), but we are asked for the time. We must compute the time separately for the periods before and after the stop. We can use the answer to (*a*) and apply Eq. (12) or (13). Evidently (12) will be simpler. We have

$$v_t = 0, \qquad v_0 = 60 \text{ mi./hr.} \qquad \text{and} \qquad a = -3{,}600 \text{ mi./hr.}^2;$$

therefore,

$$0 = 60 \text{ mi./hr.} - (3{,}600 \text{ mi./hr.}^2)t$$

or

$$\frac{60}{3{,}600} \text{ hr.} = \frac{1}{60} \text{ hr.} = 1 \text{ min.}$$

For the acceleration after the stop, we can again apply Eq. (12). We have $v_0 = 0$, $v_t = 60$ mi./hr., and $a = 900$ mi./hr.^{2}; therefore,

$$60 \text{ mi./hr.} = 0 + (900 \text{ mi./hr.}^2)t$$

$$t = \frac{60}{900} \text{ hr.} = \frac{1}{15} \text{ hr.} = 4 \text{ min.}$$

Since the train spent 1 min. slowing down, 2 min. at the station, and 4 min. speeding up again, the total elapsed time was 7 min.

PROBLEMS

1. Draw graphically and roughly to scale the displacements necessary to reach the library from the physics laboratory. Neglect vertical displacements. Measure the resultant distance to the library.

2. Resolve the vectors of Prob. 1 into components along the x- and y-axes. Compute the magnitude and direction of the resultant analytically.

3. What is the average speed of a train traveling from New York to Chicago, a distance of 1,000 mi., in 18 hr.?

4. An automobile travels with a constant speed of 50 mi./hr. What is the average speed during the first hour? What is the instantaneous speed at the end of the first half hour?

5. If the telephone poles along a railroad track are 150 m. apart and a train passes 240 poles in 15 min., what is its average velocity in meters per second?

6. If the velocity of a moving object reverses its direction in t sec., what is the average acceleration?

7. A man on one side of a river 1 km. wide rows directly for the other bank with a velocity of 1 m./sec. Having reached the other side, he rows

back with the same velocity, keeping his boat pointed toward the first bank. If he lands 200 m. below his starting point, what is the velocity of the river?

8. A man on one side of a river rows directly for the other bank with a velocity of 1 m./sec. Having reached the other side, he rows back with the same speed keeping upstream at an angle of 45° with the direction of the river. How fast was the river traveling if he landed just at his starting point?

9. A man bowls in an alley 40 ft. long by 7 ft. wide. His ball starts in the middle of the alley and goes off the side just as it reaches the other end. If it takes $1\frac{1}{2}$ sec. for the ball to go off, what is the average velocity of the ball?

10. An airplane flying south has a speed relative to the air of 120 mi./hr. If there is a wind of 15 mi./hr. blowing south, what is the speed of the plane relative to the earth?

11. If the wind in Prob. 10 is from the west, what is the speed of the plane relative to the earth?

12. An air-mail pilot flies between two cities 150 mi. apart. His ordinary flying speed is 100 mi./hr. If he is 20 min. late because he had to fly directly into the wind, what is the velocity of the wind?

13. Give some examples of the change of velocity at constant speed.

14. If a car starts from rest at a stop light and at the end of 6 sec. is going 50 mi./hr., what is its average acceleration?

15. An airplane flying at a speed of 150 mi./hr. encounters a wind of 40 mi./hr. from the southwest. In what direction must the plane head in order to travel eastward, and what will its velocity be relative to the ground?

16. If the velocity changes its direction by an angle θ in t sec., what is the average acceleration?

17. A car traveling 30 mi./hr. skids around a right-angle corner in 1 sec. What is its average acceleration, and which way is it directed?

18. A body moving in a straight line has a speed of 10 m./sec. and 7 sec. later has a speed of 45 m./sec. Find the average acceleration.

19. An arrow being shot from a bow is accelerated over a distance of 60 cm. If its initial velocity in flight is 70 m./sec., find its average acceleration as it was flung from the bow.

20. Find the average acceleration of an automobile that changes its speed from 16 km./hr. to 64 km./hr. in $\frac{1}{2}$ min.

21. If the velocity of a body changes from 45 km./hr. to 15 km./hr. in 10 sec., find its average acceleration.

22. If the train of Prob. 5 started from rest at the beginning of the 15-min. interval, what was its average acceleration?

23. A body initially at rest moves for t sec. with an acceleration a. Write an expression for the average velocity during the interval of time t.

24. A body originally at rest is given an acceleration a for a time of t sec. How far does the body travel in the t sec.?

25. If the brakes on a train are able to give it a negative acceleration of 3 m./sec./sec., find the shortest space in which the train can be stopped when traveling at the rate of 20 m./sec.

26. A uniformly accelerated sled is timed during two different seconds with a 2-sec. intermission between timings. If it went 8 m. during the first interval of timing and 72 m. during the last, find its acceleration.

27. A sled starts from the top of a hill with a speed of 2 m./sec. and experiences an acceleration of 0.75 m./sec./sec. How far does it travel during the third second?

28. A train leaving New York reaches its constant running speed of 70 mi./hr. in 4 min. What is its average acceleration? How far does it travel before reaching such a speed? If its brakes decelerate it at the same rate and it makes two other stops before reaching Philadelphia, how long does it take to cover the 90 mi. between New York and Philadelphia?

CHAPTER IV

NEWTON'S LAWS OF MOTION. GRAVITATION

Force and Motion

1. In the last chapter, we discussed the general qualitative relations between force, mass, and motion derived from ordinary experience. We are now in a position to consider these relations quantitatively. First of all, how does a moving body behave if there are no forces acting on it? The answer to this question was given clearly by Newton whose conclusions may be summed up in the following statement:

Every body remains in its condition of rest or of uniform motion in a straight line except as it is compelled to change that condition by an external force applied to it.

2. This is known as Newton's First Law of Motion and has been verified by many thousands of observations since his time. It is essentially a qualitative definition of force (see Par. 17 below for a discussion of this point). It means that there can be no change in either the direction or magnitude of the velocity of a moving body unless some force is acting on it and, conversely, that if there is an uncompensated force acting on a body the velocity of the body must be changing in either magnitude or direction or both. In other words, there is some connection between force and acceleration.

3. In order to state the precise connection which experiment has shown to exist between force and acceleration, we need to define the quantity known as momentum. The effects of a moving body depend on both its mass and its velocity, and we can express this familiar fact by speaking of the quantity of motion or the momentum of the body. In precise terms, *the momentum of a moving body is the product of its mass and velocity.* Newton expressed the relation between force and motion in terms of change of momentum, and the development of the theory of relativity has shown that this is better than to use acceleration, *i.e.*, rate of change of velocity, on the assumption that the mass is

constant. The conclusion which Newton drew from his observations and which has been verified many times since may be expressed as follows:

The rate at which the momentum of a body changes is proportional to the impressed force causing the change, and this change takes place in the direction of the straight line in which the force acts.

4. Momentum was defined as quantity of motion or more specifically as mass times velocity. It is a vector quantity. Its rate of change may be written in the same way as the rate of change of position or velocity. Thus, if the momentum of a body at the time t_0 is M_0 and at the time t is M, the average rate of change of the momentum in the interval is

$$\frac{M - M_0}{t - t_0} \propto F_{av}, \text{ the average force,} \tag{1}$$

and if the time interval is allowed to approach zero, the instantaneous rate of change of momentum and the corresponding force acting are given by

$$\frac{\Delta M}{\Delta t} \doteq \frac{dM}{dt} \propto F. \tag{2}$$

In the problems with which we shall be dealing most of the time, the masses of the moving bodies may be taken as constant. Therefore, since $M_0 = mv_0$ and $M = mv$, $\Delta M = m\Delta v$ and

$$\frac{\Delta M}{\Delta t} = m\frac{\Delta v}{\Delta t} \doteq m\frac{dv}{dt}. \tag{3}$$

But dv/dt has been defined as the acceleration. Hence, Newton's second law becomes $F \propto ma$; or, in words, force is proportional to mass times acceleration. This may be written as an equation

$$F = kma \tag{4}$$

where k is a factor of proportion whose value depends on the units used. This is a relation that is used continually in all problems of motion. It is a conclusion based on experiment.

5. We defined the unit of mass that is used in scientific work as the kilogram. We recalled the definitions of unit length and unit time already familiar. As we have seen, these three units the meter, the kilogram, and the second form the basis of a whole system of units of the various quantities with which physics

deals. This system is commonly referred to as the m.k.s. system of units. The unit of velocity in this system is the meter per second and the unit of acceleration is the meter per second per second. The unit of momentum is the kilogram-meter per second. There is no particular reason why these units should not have names of their own as the units of force and energy do. Probably the reason they do not is because it has been found that the self-explanatory nature of a term like "m. per sec.^{2}" outweighs its lack of conciseness. An interesting unit of velocity that does have a name of its own is the "knot" which is the name given to the seagoing unit of velocity, one nautical mile per hour.

6. We now wish to define the unit of force* in the m.k.s. system. It is defined as a force that changes the momentum of a moving body at the rate of one kilogram-meter per second per second and is called the newton. Repeating the definition in more elementary form: *A force of one newton is a force that will give a mass of one kilogram an acceleration of one meter per second per second.* Thus if the quantities in Eqs. (1), (2), and (3) are all expressed in m.k.s. units, the signs of proportionality may be replaced by equal signs, and k in Eq. (4) becomes unity so that the equations may be rewritten

$$F_{av} = \frac{M - M_0}{t - t_0}$$

$$F = \frac{\Delta M}{\Delta t} \doteq \frac{dM}{dt}$$

$$F = m\frac{\Delta v}{\Delta t} \doteq m\frac{dv}{dt} = ma$$

or simply

$$F = ma. \tag{5}$$

* In the c.g.s. system, the unit of force is called the dyne and is defined as the force that will give to a mass of one gram an acceleration of 1 cm./sec.^{2} One newton equals 10^5 dynes.

The desire to have k equal to one and yet express the mass in pounds and acceleration in feet per second per second leads to the introduction of a unit of force for the English system of units corresponding to the dyne in the c.g.s. system. It is called the poundal and is defined as follows: The poundal is a force that will give a mass of one pound an acceleration of 1 ft./sec.^{2} The dyne (or newton) and the poundal are often called absolute units of force.

This last relation, force equals mass times acceleration, is probably the most useful single relation in physics, and the student should learn how and when it can be applied.

7. It should be pointed out that the procedure followed in defining the newton is one that is frequently used and should be examined carefully. In general, it goes like this: We have several different quantities which we can designate by A, B, C, D, etc. By experiment and observation, it is found that there is a dependence of these quantities on each other. For example, A may increase proportionally as B increases, A may also increase proportionally as C decreases, and perhaps A decreases proportionally as the square of D increases. Such a relation would be expressed by writing the proportion

$$A \propto \frac{B}{CD^2}$$

or the equation

$$A = k\frac{B}{CD^2}$$

where k is a factor of proportion. *This relation expresses an experimental result.* If the relation is general and quantitative, it must hold when the different quantities are expressed numerically in appropriate units. The equation will be most simple if the factor of proportion k is equal to unity. If one of the quantities involved, say A, is one for which no unit has yet been defined, k can be arbitrarily set equal to unity and the equation used to define the unit of A in terms of the units of the other quantities. This is what was done in the definition of the newton. But if the units of all the quantities were already defined from other experiments, this would be impossible and k would have to be kept in the equation with the appropriate numerical value. An instance of the latter case is encountered if the kilogram weight or pound weight is taken as the unit of force.

The Law of Gravitation

8. In the work of Tycho Brahe (1546–1601), Kepler (1571–1630), and Newton (1642–1727) on the motion of the planets we have a beautiful example of the scientific method. Observations of the positions of the planets over a long period of years were made by Tycho Brahe. From these observations, Kepler

deduced three comparatively simple empirical laws. Then Newton showed that these empirical geometrical laws could all be explained in terms of one universal law, the law of gravitation, which may be stated as follows:

Every particle of matter in the physical universe attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

Or in mathematical language,

$$F = G\frac{m_1 m_2}{d^2} \tag{6}$$

where F is the force, m_1 and m_2 the two masses, d the distance between the two particles, and G the constant of proportion whose numerical value depends on the units in which the masses, distance, and force are expressed. As stated above, the law applies to any two particles of matter so small that they may each be considered concentrated at a point and the distance d measured between the two points. In the case of extended bodies, this attraction will exist between each particle in a body and every other particle in the same body and the other bodies. The calculation of the resultant force appears at first sight to be of impossible complexity. Fortunately this is far from true. Newton himself showed by a very pretty mathematical proof that a sphere of uniform density will give the same resultant effect as an equal mass concentrated at the point at the center of the sphere. Though the attractions between irregularly shaped bodies cannot be calculated so easily, good approximations can be obtained by the application of the calculus.

9. Newton deduced the law of gravitation from the observed motion of the planets. He then verified it by calculating the motion of the moon and showing that his results agreed with observations. But he had no way of measuring the masses of the sun, the planets, or even the moon and earth. Consequently he was not able to determine the value of the gravitational constant G. In order to do this, some method of observing the gravitational attraction between bodies of known masses was required. The first experiment of this kind was performed by Cavendish in the latter part of the eighteenth century.

10. Cavendish reported his experiments to the Royal Society of London on June 21, 1798, in a paper entitled "Experiments To Determine the Density of the Earth." For a description of his apparatus, we can hardly do better than to quote his own words and reproduce his own diagram. He says:

The apparatus is very simple; it consists of a wooden arm, 6 feet long, made so as to unite great strength with little weight. This arm is suspended in an horizontal position, by a slender wire 40 inches long, and to each extremity is hung a leaden ball, about 2 inches in diameter; and the whole is inclosed in a narrow wooden case, to defend it from the wind.

As no more force is required to make this arm turn round on its centre than what is necessary to twist the suspending wire, it is plain, that if the wire is sufficiently slender, the most minute force, such as the attraction of a leaden weight a few inches in diameter, will be sufficient to draw the arm sensibly aside. The weights . . . were 8 inches in diameter. One of these was to be placed on one side the case, opposite to one of the balls, and as near it as could conveniently be done, and the other on the other side, opposite to the other ball, so that the attraction of both these weights would conspire in drawing the arm aside. . . . As the force with which the balls are attracted by these weights is excessively minute, not more than 1/50,000,000 of their weight, it is plain, that a very minute disturbing force will be sufficient to destroy the success of the experiment.

Cavendish then describes the disturbing effects of air currents and temperature differences that made it desirable to put the apparatus in a closed room and make readings and adjustments from outside. He then describes the details of the apparatus as follows:

Figure 28 is a longitudinal vertical section through the instrument, and the building in which it is placed. *ABCDDCBAEFFE*, is the case: *x* and *x* are the two balls, which are suspended by the wires *hx* from the arm *ghmh* which is itself suspended by the slender wire *gl*. This arm consists of a slender deal rod *hmh* strengthened by a silver wire *hgh*; by which means it is made strong enough to support the balls, though very light.

The case is supported, and set horizontal, by four screws, resting on posts fixed firmly into the ground: two of them are represented in the figure, by *S* and *S*; the two others are not represented, to avoid confusion. *G* and *G* are the end walls of the building. *W* and *W* are the leaden weights; which are suspended by the copper rods *RrPrR*, and the

wooden bar *rr*, from the center pin *Pp*. This pin passes through a hole in the beam *HH*, perpendicularly over the centre of the instrument, and turns round in it, being prevented from falling by the plate *p*. *MM* is a pulley, fastened to this pin; and *Mm*, a cord wound round the

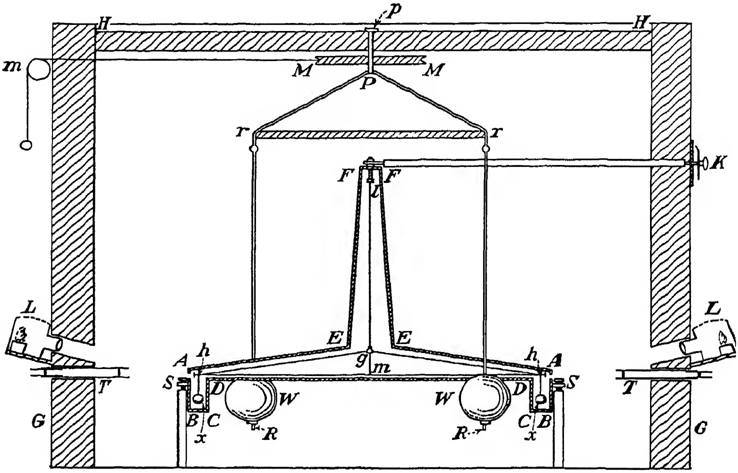


FIG. 28.—Cavendish's experiment (elevation). (*After Cavendish.*)

pulley, and passing through the end wall; by which the observer may turn it round, and thereby move the weights from one situation to the other.

Figure 29 is a plan of the instrument. *AAAA* is the case; *SSSS*, the four screws for supporting it. *hh*, the arm and balls. *W* and *W*,

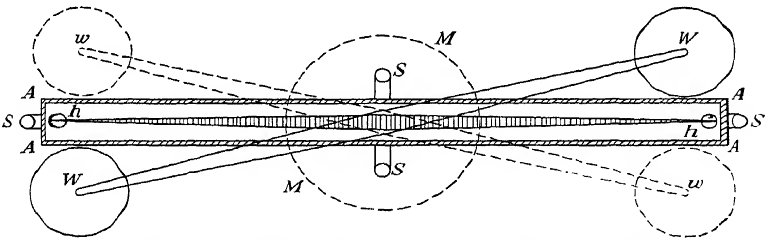


FIG. 29.—Cavendish's experiment (plan). (*After Cavendish.*)

the weights. *MM*, the pulley for moving them When the weights are in this position, both conspire in drawing the arm in the direction *hW*; but, when they are removed to the situation *w* and *w*, represented by

the dotted lines, both conspire in drawing the arm in the contrary direction *hw*. These weights are prevented from striking the instrument, by pieces of wood, which stop them as soon as they come within $\frac{1}{5}$ of an inch of the case. The pieces of wood are fastened to the wall of the building; and I find, that the weights may strike against them with considerable force, without sensibly shaking the instrument.

In order to determine the situation of the arm, slips of ivory are placed within the case, as near to each end of the arm as can be done without danger of touching it, and are divided to 20ths of an inch. Another small slip of ivory is placed at each end of the arm, serving as a vernier, and subdividing these divisions into 5 parts; so that the position of the arm may be observed with ease to 100ths of an inch, and may be estimated to less. These divisions are viewed, by means of the short telescopes T and T, (Fig. 28) through slits cut in the end of the case, and stopped with glass; they are enlightened by the lamps L and L, with convex glasses, placed so as to throw the light on the divisions; no other light being admitted into the room.

The divisions on the slips of ivory run in the direction Ww, (Fig. 29) so that, when the weights are placed in the positions w and w, represented by the dotted circles, the arm is drawn aside, in such direction as to make the index point to a higher number on the slips of ivory; for which reason, I call this the positive position of the weights.

FK (Fig. 28) is a wooden rod, which, by means of an endless screw, turns round the support to which the wire gl is fastened, and thereby enables the observer to turn round the wire, till the arm settles in the middle of the case, without danger of touching either side. The wire gl is fastened to its support at the top, and to the centre of the arm at the bottom, by brass clips, in which it is pinched by screws.

In these two figures, the different parts are drawn nearly in the proper proportion to each other, and on a scale of 1 to 35.

The details of procedure and measurement as reported by Cavendish are too long and involved to be suitable for quotation. In principle, what he did was to observe the deflection of the arm hh when the large weights were in the "positive" position and again when they were in the "negative" position. The average of these then gave him the amount that the suspending wire gl was twisted by the force of attraction between the large weights W and the small lead balls x. In order to deduce from this observation the force of attraction producing the twist, it was necessary to determine the elastic constants of the suspending wire. If the modulus of rigidity of the wire (see Chap. II, Par. 32) had been known exactly, this might have been cal-

culated, or it might perhaps have been determined experimentally by producing the same twist by actually applying measurable forces to the ends of the arm, by threads leading over pulleys to tiny weights, for example. In fact, Cavendish used a much better method that we cannot fully understand until we have studied rotational and periodic motion. He determined the natural period of oscillation of the moving system and from this deduced the force necessary to twist the system a given amount (Chap. XXIII, Par. 17). To be specific, he found that the force required to "draw the arm one division out of its natural position is $\frac{1}{818}N^2$ of the weight of the ball" where N is the period of oscillation of the arm measured in seconds. Therefore, by using the observed deflection caused by the weights W, he could express the force between them and the lead balls in terms of the weight of one ball. This came out to be about 1/48,000,000 the weight of a ball. Cavendish was attempting to determine the density of earth not the constant of gravitation G. The earth's density is deduced from his data in the following way:

Express the weight of a body, say one of the lead balls on the arm of Cavendish's balance, in terms of the gravitational law of Par. 8. This gives us

$$\text{Weight of ball} = F_1 = G\frac{m_b m_e}{R^2} \qquad (7)$$

where G is the gravitational constant, m_b is the mass of the ball, m_e the mass of the earth, and R is the distance from the center of the earth to the center of the ball, *i.e.*, the earth's radius. Similarly the attraction between a ball and one of the weights W will be given by the equation

$$F_2 = G\frac{m_b m_w}{r^2} \qquad (8)$$

where m_w is the mass of the weight W and r is the distance between the centers of the weight and the ball. Dividing F_2 by F_1, we get

$$\frac{F_2}{F_1} = \frac{m_w R^2}{m_e r^2} \qquad (9)$$

The left-hand side of this equation is the *ratio* measured in Cavendish's experiment. The r and m_w are also measured in his experiment, and R, the radius of the earth, is known. There-

fore the only remaining quantity, the mass of the earth, is determined. If the mass of the earth and its volume are known, its density can be determined. The value deduced by Cavendish was 5.48 times the density of water. The accepted value of this ratio at the present time is 5.52 corresponding to a value of approximately 5.95×10^{24} kg. for the earth's mass. To get the value of the gravitational constant G, additional information is needed. We must substitute the mass of the earth, the radius of the earth, and the mass of the ball on the right-hand side of Eq. (7), and we must measure the force F_1, the weight of the ball, not merely the ratio of F_2 to F_1. This might be measured by a spring balance calibrated directly in newtons; or, since we have defined force in terms of the acceleration it produces, if we observe the acceleration of the ball when it is acted on by no other force than its own weight, we can compute that weight in newtons. This is a very simple experiment and is usually discussed before anything is said about the law of gravitation or the value of the constant G. Nevertheless we prefer to omit further discussion of it in order to avoid a certain difficulty that usually causes many students trouble. At present, we shall merely quote the value of G obtained by this process.

11. The value of G calculated from Cavendish's data is 6.717×10^{-11} newton-m.^{2}/kg.^{2} As the result of more accurate repetitions of his experiment, the value of G is now thought to be

$$6.66 \times 10^{-11} \text{ newton-m.}^2/\text{kg.}^2$$

That is if the masses m_1 and m_2 in Eq. (6) are expressed in kilograms and the distance d in meters, then the force F between them will be correctly given in newtons if the preceding number is substituted for G. Phrasing it in a slightly different way, two small spheres, each 1 kg. in mass and with their centers 1 m. apart, will attract each other with a force of 6.66×10^{-11} newton. It is obvious that so small a force is very difficult to measure and that gravitational forces become of importance only when at least one of the masses involved is very large. Thus on the earth's surface the mutual gravitational attractions between molecules, motor cars, and battleships are negligibly small compared with the other forces acting on them. But the forces of gravitational attraction between these bodies and the earth itself are of a very different order of magnitude.

The Earth's Gravitational Attraction

12. As we have seen, the mass of the earth is very nearly 5.95×10^{24} kg. The radius of the earth is found to be approximately 6.36×10^6 m. This latter number is so large compared with the variations in the distance from the earth's center caused by different latitudes and elevations above sea level that such variations may be neglected in our present calculations. Therefore if the earth is assumed to be a sphere of uniform density, the force of gravitational attraction toward the earth's center on any body of mass m kg. on the earth's surface is

$$F = 6.66 \times 10^{-11} \frac{5.95 \times 10^{24} m}{(6.36 \times 10^6)^2} = 9.8\, m \text{ newtons.} \tag{10}$$

Thus the earth's gravitational attraction for any body is 9.8 newtons/kg. This quantity is universally designated by g and the gravitational pull of the earth on a body of mass m is mg. Consequently, a freely falling mass of 1 kg. will move toward the center of the earth with an acceleration of 9.8 m./sec.^{2} If the mass of the body is 5 kg., the force acting on it is 49 newtons and the acceleration is $a = F/m = \frac{49}{5} = 9.8$ m./sec.^{2} again. This calculation is hardly necessary, for we see that the gravitational pull is always proportional to the mass and the acceleration is always inversely proportional to the mass and directly proportional to the force so that the acceleration will be the same for all masses. In mathematical language,

$$F = G\frac{M}{R^2}m \qquad \text{and} \qquad a = \frac{F}{m};$$

therefore

$$a = G\frac{M}{R^2} = g \tag{11}$$

where M is the mass of the earth, R its radius, and G the gravitational constant. Therefore the acceleration of falling bodies is independent of their mass.

13. We calculated the figure of 9.8 newtons for the force of gravity on 1 kg. by taking an approximate value of the earth's radius. Since the earth is somewhat flattened at the poles, the gravitational attraction is somewhat greater there than at the equator. Furthermore, there is another force acting on bodies on the earth's surface, namely, the centrifugal force resulting from

the earth's rotation. This is very small even at the equator and is zero, of course, at the poles. But it is sufficient to lessen the resultant acceleration of a falling body. Experimental determinations show that these two effects cause g to vary from 9.77989 m./sec.^{2} at the equator to 9.83210 m./sec.^{2} at the pole.* If we imagine ourselves transported to the surface of the moon, we experience a much greater variation of the gravitational attraction. The gravitational force on 1 kg. on the moon's surface is about one-sixth as much as on the earth, or about 1.67 newtons.

14. The force of gravitational attraction on a body on the earth's surface is called the weight of that body. The weights of bodies are proportional to their respective masses. Although the factor of proportion evidently varies from place to place on the earth's surface, it is the same for all bodies at a given place. Consequently, the masses of different bodies can be compared by weighing at the same place. Gravity never ceases to act and is unaffected by temperature, change of state, chemical reaction, or any other known influence. Therefore the weights of bodies are the most familiar of all forces. For the purposes of this book, we can assume that the weight of m kg. of matter is $9.8 \times m$ newtons or, in the English system, that the force exerted by gravity on a mass of m lb. is $32 \times m$ poundals. Since at any given point, the weight of 1 kg. or of 1 lb. is a force of definite amount, it is possible to use these weights as units of force, but it must be pointed out that in that case the value of k in Eq. (4) is not unity. If the mass is in kilograms, the acceleration in meters per second per second, and the force in kilograms weight, the value of k is 1/9.8. If the mass is in pounds, the acceleration in feet per second per second, and the force in pounds weight, the value of k is 1/32. It would be perfectly possible to find the value of k that would allow one to express the mass in kilograms, the acceleration in miles per hour per minute, and the force in poundals and still have the equation $F = kma$ valid. But experience shows that the best procedure for handling the problems of physics and engineering is always to use a system of units in which k is unity, such as newtons, kilograms, and meters per

* The determination of g is usually carried out in several ways in a school laboratory. It can be done by direct observation of a falling body or by pendulum measurements (see Chap. XXIII).

second per second, or poundals, pounds, and feet per second per second. Then the equation is simply $F = ma$. For convenience in reducing forces to these units, it is to be remembered that

One pound weight = 32 poundals

and

One kilogram weight = 9.8 newtons

or, symbolically,

$$W = mg \tag{12}$$

where g is the acceleration of a freely falling body on the earth's surface.

Action and Reaction

15. It has just been pointed out that the force of gravity never ceases to act. Also it follows from Newton's second law of motion that any body acted on by an uncompensated force will be accelerated. How is it, then, that most objects on the earth's surface are at rest? The total force acting on them must be zero. Consequently there must be forces acting on them equal in magnitude and opposite in direction to their weights. Thus, the table must be pushing up on this typewriter with a force just equal to the downward pull of the weight of the typewriter. A little consideration shows that this idea is in harmony with ordinary experience. A man sitting in a chair pushes down on the seat of the chair, but the chair is pushing up at the same time on the man. A tug of war may be so evenly matched that the rope is not moving though each team may be pulling for all it is worth. Even when there is motion, the inertial reaction of the accelerated body pushes back on the source of the accelerating force. The earth exerts a pull on a freely falling body, but so does the body exert an equal pull on the earth. Thus all forces occur in pairs. Every force acting on a body is accompanied by an equal and oppositely directed force acting on some other body. These observations were described by Newton in his third law of motion which may be stated as follows:

To every action there is an equal and opposite reaction.

16. Newton's three laws of motion are no mere abstract generalizations from experience. They are practical principles applicable to the solution of problems of mechanics. Because of their great importance, we shall repeat them.

Newton's First Law. **Every body remains in its condition of rest or of uniform motion in a straight line except as it is compelled to change that condition by an external force applied to it.**

Newton's Second Law. **The rate at which the momentum of a body changes is proportional to the impressed force causing the change; and this change takes place in the direction of the straight line in which the force acts.**

Newton's Third Law. **To every action there is an equal and opposite reaction.**

17. We have introduced these laws as if they were observed relations between concepts clearly defined by other observations. This is too great a simplification. Newton's laws are by no means simple statements of observation like the statement that night follows day. They start from intuitive concepts of position, time, mass, and force. Precise definitions are then set up for position and mass which are as nearly as possible equivalent to our intuitive notions. Other quantities like velocity, acceleration, and momentum are introduced and defined. These are slightly less familiar even in vague form. Once these definitions are clearly set out, we proceed to make observations on the positions, velocities, and accelerations of various bodies under various circumstances. We find that we can talk about these observations accurately and consistently if we use our definitions and Newton's laws of motion. It is not until we have introduced Newton's first and second laws that we are able to give a precise definition of what we mean in physics by a force. Thus, in a sense, Newton's first law is a postulate, a qualitative definition of force, and the second law is a quantitative definition of force. But these definitions are consistent with the crude notions of force that we already have from ordinary experience in which we identify force and muscular effort. The third law also conforms to common experience. In kicking a football, there is certainly a force exerted on the foot by the ball as well as on the ball by the foot. The first law may seem contrary to observation if we neglect to introduce the concept of friction. We cannot in practice ever get a body to move with constant velocity without applying a force. We explain this by saying that there is always some friction acting to retard the motion. This frictional retarding force may be made small and can be interpreted in terms of intermolecular forces. Since we can

interpret and control it almost to the point of elimination, we prefer to recognize its presence and accept the validity of Newton's laws. Thus, when we say that no exception to Newton's laws has ever been noted, we mean that the language set up by Newton's laws and by the previous definitions is adequate to describe every situation that has been observed.

Equilibrium

18. In discussing the third law, the case of a body's remaining motionless under the action of several forces has been mentioned. We now wish to investigate what conditions must be satisfied by the forces if the body is not to move (or more strictly is not to accelerate). In technical language, what are the conditions for equilibrium? Two cases may be distinguished, that of a particle and that of an extended body. In the case of a particle, the forces must all be acting at the same point and the only motion that can occur is one of translation.* The particle is so small that it is considered to be concentrated at a point so that motion of rotation has no meaning. For such a particle, it is fairly obvious that the forces acting can be added vectorially and that if their resultant is zero the particle will be in equilibrium. Thus for a particle the condition for equilibrium is that the vector sum of all the forces acting on the particle shall be zero.

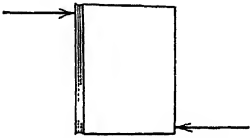


FIG. 30.—Turning effect of a couple.

19. For an extended body the situation is different. Think of a book lying on a smooth table. Push the book equally hard in opposite directions as indicated in Fig. 30. The vector sum of the two forces is zero, and the center of the book does not move but the book turns. Now keep the directions and magnitudes of the pushes the same, but move their points of application in toward the center of the book. It is found that the book has less and less tendency to turn and that when the forces are applied along the same line the book does not turn at all. Consider the

* A motion of pure translation for an extended body is a motion such that the displacement of every point in the body is equal and parallel to the displacement of every other point. A motion of pure rotation is one where one line in the body remains motionless and the other points in the body move in circles around that line as an axis. Any motion can be obtained by combinations of pure rotations and pure translations.

case of a seesaw as illustrated in Fig. 31. There are four forces acting on the board in this figure, the weight of the plank which may be considered as acting at its center, the weights at A and at B, and the reaction of the support at D. All these forces are vertical. Their vector sum is therefore merely their algebraic sum, and we have

$$m_A g + m_B g + mg - R = 0 \tag{13}$$

if the board is in equilibrium. By moving the weights in and out along the board, it is found that the turning effect of the weights is proportional to their distance from the support D as well as to their size. A small weight at the end of the board has the same turning effect as twice as great a weight half way

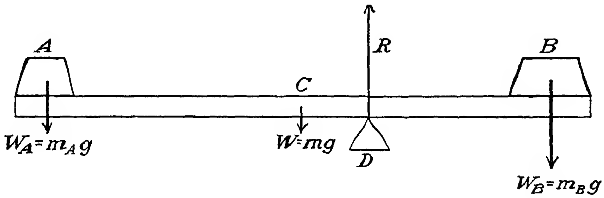


FIG. 31.—Moments of force.

between the support and the end of the board. In general, we find that the *turning effect of a force about a pivot is equal to the product of the magnitude of the force and the perpendicular distance from the pivot to the line of application of the force. This is called the moment of the force about the pivot.* It is found that if the board in our example is in equilibrium the sum of the moments of force tending to turn the board clockwise is just equal to the sum of the moments of force tending to turn the board counterclockwise. If we use the usual convention that counterclockwise rotation is positive and clockwise rotation negative, the moments can be added algebraically.

20. It is quite easy to show mathematically that if the vector sum of a system of coplanar forces acting on an extended body is zero and the sum of the moments of the forces about one point in their plane is zero, then the sum of the moments of force about every point in this plane is zero whether the point is in the body or not. Since this is true, the conditions for equilibrium of a rigid body can be stated as follows:

A body is in equilibrium if (1) *the vector sum of the forces acting on it is zero and* (2) *the sum of the moments of the forces about any axis is zero.*

Center of Mass

21. There is a particular case of static equilibrium that leads to the definition of an important point, the center of mass of an extended body. Consider any extended body hanging from a string as shown in Fig. 32(a). The body is not moving, and the forces on it must be in equilibrium. The only forces acting are the tension of the string vertically upward and the weights of the various parts of the body vertically downward. The tension

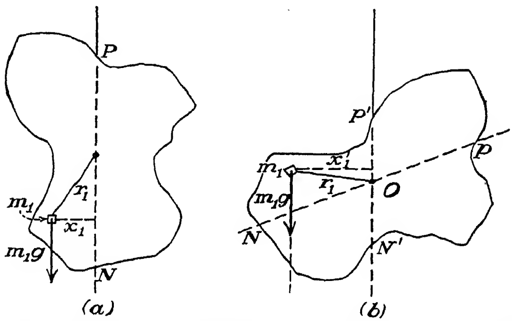


FIG. 32.—Experimental determination of the center of mass.

of the string acts vertically along the line NP passing through the point P where the string is attached. The weights of the various parts of the body act vertically along lines passing through them. For the sake of simplicity, we shall assume the body a thin sheet, a two-dimensional body. Take any representative small part of the body of mass m_1, its weight m_1g acts along a line distant x_1 from NP as shown in the diagram. This force tends to turn the body in a counterclockwise direction as do the weights of all of the parts of the body to the left of PN. This effect must be just counteracted by the tendency of the parts of the body on the right of PN to turn it in a clockwise direction. Expressing this condition mathematically by taking moments of force about P, we say that the sum of all the moments of force about P must be zero. If we designate the masses of various parts of the body by m_1, m_2, . . . , etc., and their distances from PN by

$x_1, x_2, \ldots$, etc., and call x positive if it is to the right and negative if it is to the left of PN; then the condition for equilibrium is that

$$g(m_1x_1 + m_2x_2 + m_3x_3 + \cdots) = g\Sigma m_ix_i = 0. \qquad (14)$$

22. Now shift the point where the string is attached to P'. The body will swing into a new position as shown in Fig. 32(*b*). The same argument as before can be applied, and the new condition for equilibrium is that

$$g(m_1x'_1 + m_2x'_2 + \cdots) = g\Sigma m_ix'_i = 0 \qquad (15)$$

where $x'_1, x'_2, \ldots$, etc., are the distances of the various parts of the body from the new vertical line of the supporting force $P'N'$. Since we are dealing with a two-dimensional body, the lines PN and $P'N'$ in the body must intersect. Call the point of intersection O. We shall find that to whatever point the string is attached the extension of its line through the body will pass through this same point O. This point is called the center of mass and is characterized by the condition that the vector sum of the products of each little element of mass of the body and the vector drawn to that element from O is zero. Written in mathematical form, this condition is

$$m_1\bar{r}_1 + m_2\bar{r}_2 + m_3\bar{r}_3 + \cdots = \Sigma m_i\bar{r}_i = 0 \qquad (16)$$

where the $\bar{r}$'s are vectors and this is a vector equation. If the vector sum of all these vectors is zero, the algebraic sum of their components along any direction must be zero. Furthermore, there is no reason to restrict ourselves to two dimensions. Taking any set of mutually perpendicular axes with its origin at the center of mass, we have the three equations

$$\Sigma m_ix_i = 0 \qquad \Sigma m_iy_i = 0 \qquad \Sigma m_iz_i = 0. \qquad (17)$$

It is evident therefore that the moments of gravitational force will always add up to zero about a horizontal axis in a vertical plane containing the center of mass.

23. The situation we have been considering is that of equilibrium between a single force acting through the center of mass and a system of parallel forces acting throughout the body and proportional to the masses distributed through the body. Let us suppose now that the body is resting on a frictionless horizontal

plane and that a force is applied whose line of action passes through the center of mass. Every particle in the body will exert an inertial reaction against acceleration proportional to its mass. These reactions will act just as the weights did in our previous problem; they are in effect a system of forces parallel and opposite to the external force. Since this force is acting through the center of mass, the resultant turning effect of these inertial reactions will be zero and the body will move in response to the external force as if it were a particle with all its mass concentrated at the center of mass. If the external force were not acting on a line through the center of mass, the body would turn. This general case is beyond the scope of our treatment. The special case of rotation about an axis which is fixed will be taken up in Chap. XVIII.

Conservation of Momentum

24. According to Newton's second law, the momentum of a moving system remains unchanged if there is no external force

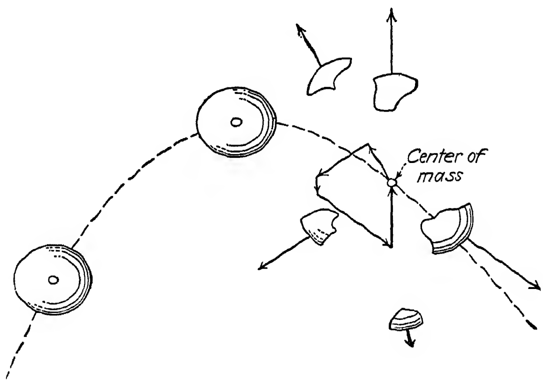


FIG. 33.—Conservation of momentum. The center of mass of the shell moves as though no explosion had taken place. The sum of the momenta of the pieces due to the explosion is zero.

acting on it. Consider an exploding projectile. Before it explodes, it has a certain momentum $M = mv$; after the explosion, each of its fragments has a certain momentum. The velocities of the fragments have varying magnitudes and directions. The masses have varying magnitudes, though their sum must equal

the mass of the original projectile if the mass of the explosive is neglected. During the explosion, there has been no external force acting, therefore the momentum has not been changed. Therefore the vector sum of the momenta after the explosion must equal the momentum of the system before the explosion. Similarly, if two billiard balls collide, they may be considered as a moving system on which no external force is acting and whose total momentum therefore remains constant, the vector sum of the momenta of the balls after the collision being equal to the vector sum of the momenta before collision. As we sometimes say, the momentum is conserved. Though these instances and others like them are direct deductions from Newton's second law, they are of sufficient importance to justify another general statement of principle in the following form:

The total resultant momentum of any group of bodies is unchanged by actions between bodies within the group.

Continuous Forces and Impulsive Forces

25. In Par. 6, we defined force in terms of rate of change of momentum. First we expressed the average force F_{av} as the ratio of the change in momentum produced by the force to the time required to produce that change, and then we went on to express an instantaneous force in terms of the acceleration at the same instant, *i.e.*, we derived the equation $F = ma$. This equation is of the greatest value and is used in problems where a body moves through a considerable distance under the action of a force that is constant or that is varying according to some known rule. Such a force may be called a continuous force. But we often encounter sudden forces which act for a very short time, so short that the position of the body on which they act does not alter appreciably while the force is being applied. Such a force is called an impulsive force.

26. A bouncing ball illustrates both continuous and impulsive forces. Between impacts, the ball is acted on only by the force of gravity which acts continuously to increase the downward momentum of the ball, *i.e.*, produces a constant downward acceleration. Every time the ball strikes, there is a sudden upward impulsive force on it which acts only during the short time that the ball is in contact with the ground. The position of the ball does not change appreciably during this short interval

of time. But its momentum changes very greatly, almost by a factor of two if the ball is very elastic, since after striking, the momentum upward is almost as great as its previous downward momentum. (Since momentum is a vector quantity, this represents a change of almost twice the original momentum.) Evidently the impulsive force must be very great compared with the weight of the ball, since it accomplishes in the time of contact a change of momentum almost equal to that brought about by the weight in the interval between contacts.

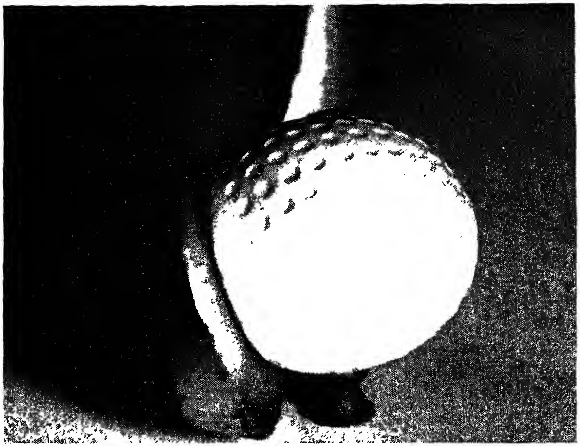


FIG. 34.—Impact of a golf club and ball illustrating large momentary magnitude of an impulsive force.

27. Problems involving impulsive forces can usually be solved by the application of Eq. (1) or by the use of the principle of the conservation of momentum (see the treatment of collisions at the end of the next chapter). However it is sometimes convenient to use the quantity called the "impulse" which is defined as the product of the average impulsive force and the time during which it acts. From Eq. (1) or the first form of Eq. (5), it is evident that

$$\text{Impulse} = F_{av}t = M - M_0 = \text{change in momentum} \quad (18)$$

where F_{av} is the average force acting in the time interval t during which the momentum of the body is changed from M_0 to M but

its position is not appreciably altered. Impulse is force multiplied by time and therefore is measured in newton-seconds in the m.k.s. system of units.

SUMMARY

According to Newton's first law of motion, the velocity of a body is constant if there is no force acting on it. Momentum or quantity of motion is the product of the mass of a body and its velocity. Newton's second law of motion states that rate of change of momentum is proportional to applied force. Usually this law is equivalent to the equation $F = ma$. The unit of force in the m.k.s. system of units is the newton, the force which gives an acceleration of 1 m./sec.^{2} to a mass of 1 kg.

The law of gravitation states that there is an attraction between any two masses given by the equation $F = Gm_1m_2/d^2$. This law explains the motion of the planets and was confirmed experimentally by Cavendish. His experiment is described in detail. On the earth's surface, the weight of any body is proportional to its mass, *i.e.*, is 9.8 newtons/kg. and the acceleration of all freely falling bodies is the same, namely, 9.8 m./sec.^{2}

Newton's third law states that to every action there is an equal and opposite reaction. Newton's three laws form the basis of mechanics and can be applied to most of its problems.

The turning effect of a force depends on its moment, which is the product of the force and the perpendicular distance from its line of action to the point about which moments are being taken. The conditions for equilibrium of an extended body are that the vector sum of the forces and the sum of the moments of force about any point must both be zero. The center of mass is defined.

The principle of conservation of momentum states that the total momentum of any system of bodies is constant if there is no outside force acting. Impulse is the product of force and time.

ILLUSTRATIVE PROBLEMS

1. A gun on a level plane fires a shell with a muzzle velocity of 600 m./sec. The angle θ which the gun barrel makes with the horizontal is 40°.

a. How long does it take the shell to reach its maximum height above the plane?

b. What is the maximum height to which the shell rises above the plane?

c. What is the total time of flight of the shell before it strikes the plane?

d. How far from the gun does the shell strike the plane?

e. At what angle does it strike the plane?

The principal point about problems of this sort is that the vertical and horizontal motions must be treated separately. If air resistance is neglected, there is no force acting in a horizontal direction and the only vertical force is that due to gravity. Consequently, the projectile moves horizontally with uniform velocity and vertically like a freely falling body, *i.e.*, with an acceleration of 9.8 m./sec.² or 32 ft./sec.² toward the center of the earth.

Take the gun muzzle as the origin of a system of coordinates, with the positive direction of y vertically upward and the positive direction of x horizontally in the direction of the target.

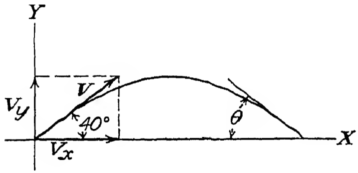


FIG. 35.—Initial velocity and range of a projectile.

a. Consider first the vertical motion. There is a positive initial velocity of $v_y = v \sin \theta = (600 \text{ m./sec.})(\sin 40°)$ upward and a negative acceleration of 9.8 m./sec.² downward. The vertical velocity will be zero when the shell is at its maximum height above the plane. The time for the negative acceleration of 9.8 m./sec.² to reduce the positive velocity of (600 m./sec.) (sin 40°) to zero is given by Eq. (12), page 65, $v_t = v_0 + at$

$$0 = (600 \text{ m./sec.})(\sin 40°) - (9.8 \text{ m./sec.}^2)t$$

or

$$t = \frac{(600 \text{ m./sec.})(\sin 40°)}{9.8 \text{ m./sec.}^2} = \frac{600 \times 0.643}{9.8} \text{ sec.} = \frac{386}{9.8} \text{ sec.} = 39.4 \text{ sec.}$$

b. The maximum height to which the shell rises is obtained directly by substituting the average velocity in terms of the initial and final velocity in Eq. (11′), page 64, thus:

$$\frac{v_0 + v_t}{2}t = \frac{(600 \text{ m./sec.})(\sin 40°) + 0}{2} 39.4 \text{ sec.}$$

$$600 \text{ m./sec.} \times 0.643 \times 39.4 \text{ sec.} \qquad 7{,}600 \text{ m.} = 7.6 \text{ km.}$$

c. The total time of flight of the shell is the time up to the highest point plus the time down again. These two times are equal. This may be verified by solving Eq. (13), page 65, for the time, since the vertical motion is one of free fall starting at rest. Thus,

$$s = v_0 t + \tfrac{1}{2}gt^2$$

$$7{,}600 \text{ m.} = 0 \text{ m./sec.} \times t + \frac{1}{2} 9.8 \text{ m./sec.}^2 \times t^2$$

$$\sqrt{\frac{2 \times 7{,}600 \text{ m.}}{9.8 \text{ m./sec.}^2}} = 39.4 \text{ sec. down.}$$

Therefore the total time of flight up plus down is

$$2 \times 39.4 \text{ sec.} = 78.8 \text{ sec.}$$

d. To get the horizontal distance the projectile travels, we have only to multiply the horizontal velocity by the time.

$$\begin{aligned} x &= v \cos 40° \times t = 600 \text{ m./sec.} \times 0.766 \times 78.8 \text{ sec.} \\ &= 460 \times 78.8 \text{ m.} = 36{,}200 \text{ m.} = 36.2 \text{ km.} \end{aligned}$$

e. The angle at which the projectile strikes is the same as the initial angle. To show this, we must get the direction of the velocity at the end of the path. The vertical component of the velocity is the same as it was initially. This may be verified from Eq. (12), page 65. Starting at rest at the top, we obtain

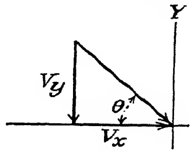


FIG. 36.—Final velocity of a projectile.

$$\begin{aligned} v_t &= v_0 + at \text{ in the form} \\ v_y &= 0 - 9.8 \text{ m./sec.}^2 \times 39.4 \text{ sec.} = -386 \text{ m./sec.} \end{aligned}$$

The horizontal velocity remains constant at 460 m./sec. Therefore the tangent of the angle θ at which the projectile strikes is given by

$$\tan \theta = \frac{v_y}{v_x}$$

or

$$\begin{aligned} \theta &= \arctan \frac{v_y}{v_x} = \arctan \frac{386 \text{ m./sec.}}{460 \text{ m./sec.}} \\ &= \arctan 0.839 = 40°. \end{aligned}$$

2. Now suppose that the gun of Prob. 1 were located 200 m. above a level plane. What are the answers to Prob. 1?

a. The maximum height above the plane is reached in the same time as in Prob. 1, *i.e.*, 39.4 sec.

b. The shell rises the same distance above the gun or 7.60 km. which is 7.60 km. + 0.2 km. = 7.80 km. above the plane.

c. The total time of flight is as before, the time up plus the time down. The time down is obtained as before:

$$s = v_0 t + \tfrac{1}{2} g t^2$$

$$7{,}800 \text{ m.} = 0 \text{ m./sec.}\, t + \frac{1}{2} 9.8 \text{ m./sec.}^2\, t^2$$

$$t = \sqrt{\frac{2 \times 7{,}800 \text{ m.}}{9.8 \text{ m./sec.}^2}} = 39.9 \text{ sec. down.}$$

Therefore the total time of flight up plus down is

$$39.4 \text{ sec.} + 39.9 \text{ sec.} = 79.3 \text{ sec.}$$

d. Again to get the horizontal distance the projectile travels, we have only to multiply the horizontal velocity by the time.

$$x = v \cos 40°t = 600 \text{ m./sec.} \times 0.766 \times 79.3 \text{ sec.}$$
$$= 460 \times 79.3 \text{ m.} = 36{,}500 \text{ m.} = 36.5 \text{ km.}$$

e. Here again, we must get the direction of the velocity at the end of the path. The vertical component of the velocity can be obtained by applying Eq. (12), page 65, $v_t = v_0 + at$ in the form

$$v_y = 0 - 9.8 \text{ m./sec.}^2\ 39.9 \text{ sec.} = -391 \text{ m./sec.}$$

The horizontal velocity remains constant at 460 m./sec. Therefore the tangent of the angle θ, Fig. 36, at which the projectile strikes is given by

$$\tan\theta = \frac{v_y}{v_x}$$

or

$$\theta = \arctan\frac{v_y}{v_x} = \arctan\frac{391}{460}$$
$$= \arctan 0.850 = 40°22'.$$

3. Suppose a car of mass M runs on a track making an angle θ with the horizontal, Fig. 37. The car is acted on by a constant force T parallel to the plane. Neglecting friction, what is the acceleration? Consider the forces on the system. There is the weight Mg of the mass of the car, acting vertically, and the reaction of the track on the car perpendicular to the direction of motion, serving merely to keep the system in its path. Finally there is the force T parallel to and up the plane.

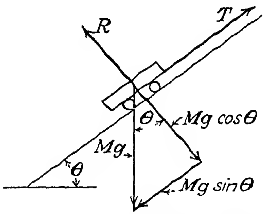


FIG. 37.—Forces acting on a car on an inclined plane.

Resolve Mg into components along the plane and perpendicular to it. The component $Mg \cos\theta$ perpendicular to the track is just counteracted by the reaction of the track. The other component $Mg \sin\theta$ tends to move the car down the track. T tends to move the car up the track. If we call motion up the plane positive, then the force acting is

$$T - Mg \sin\theta.$$

The mass of the moving system is M so that, since from Eq. (5), page 74,

$$F = ma$$
$$T - Mg \sin\theta = Ma$$
$$a = \frac{T - Mg\sin\theta}{M}$$

where T must be expressed in absolute units of force, newtons for instance, if M is in kilograms, $g = 9.8$ m./sec.^{2}, and a in meters per second per second.

4. *Atwood's Machine.* A string runs over a pulley, Fig. 38, and has a mass m_1 on one end and m_2 on the other end. The mass of the string and the mass and friction of the pulley may be neglected. Find the acceleration of the system and the tension in the string. Assume that $m_1 > m_2$ so that m_1 will fall and m_2 will rise. The net force producing acceleration in the system is the difference between the attraction of the earth for the masses m_1 and m_2, or

$$m_1g - m_2g.$$

Since the mass of the moving system includes both m_1 and m_2, we have from Eq. (5), page 74,

$$F = ma$$

$$m_1g - m_2g = (m_1 + m_2)a;$$

therefore,

$$a = \frac{m_1 - m_2}{m_1 + m_2}g.$$

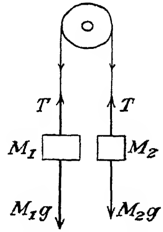


FIG. 38.—Atwood's machine.

To arrive at this result, we have considered both the masses as one moving system. To get the tension, T, in the string, we shall consider them as separate systems but as having the same acceleration. Then the force on m_1 acting down is

$$m_1g - T;$$

therefore,

$$a = \frac{m_1g - T}{m_1}.$$

On m_2, the force up is

$$T - m_2g;$$

therefore,

$$a = \frac{T - m_2g}{m_2}.$$

Since the string is taut and weightless, the tensions as well as the accelerations are the same. Consequently,

$$\frac{m_1g - T}{m_1} = \frac{T - m_2g}{m_2}$$

or

$$m_1m_2g - Tm_2 = Tm_1 - m_1m_2g;$$

therefore,

$$T = \frac{2m_1m_2g}{m_1 + m_2}.$$

Had we eliminated T from the two equations for a and solved for a, we would obtain

$$a = \frac{m_1 - m_2}{m_1 + m_2}g$$

as before, showing that this is a legitimate way of approaching the problem.

5. A ladder of length L and mass M rests against a smooth vertical wall with its foot a distance d from the foot of the wall. A man of mass m is one-third of the way up the ladder. What are the direction and magnitude of the reaction of the ground against the ladder?

This is a problem in equilibrium. The ladder does not move, therefore there must be no resultant force or moment of force on it. Draw a diagram showing the forces. Call the reaction of the wall against the ladder P. Since the wall is perfectly smooth, this force must be perpendicular to it. Call the unknown reaction of the ground R and the angle it makes with the ground φ.

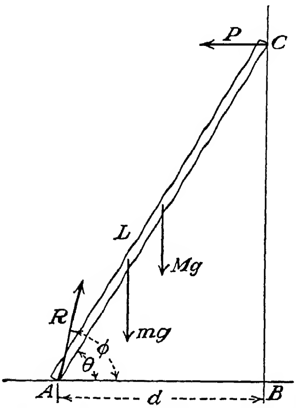


FIG. 39.—A ladder leaning against a smooth wall.

From the conditions of equilibrium, the vector sum of the forces acting on the ladder must be zero and the sum of the moments of the forces about any axis must be zero. The first condition can be satisfied by making the sum of the vertical components of the forces on the ladder equal to zero and the sum of the horizontal components of the forces equal to zero, or what amounts to the same thing, the forces up equal the forces down and the horizontal forces to the left equal those to the right.

The vertical forces are

$$R \sin \varphi = Mg + mg. \tag{19}$$

The horizontal forces are

$$R \cos \varphi = P. \tag{20}$$

To satisfy the second condition, we pick any axis and set the moments about it equal to zero. It is often convenient to pick the axis through a point at which unknown forces act, as they are then eliminated from the equation. For this reason, we choose the point A and take moments about an axis perpendicular to the plane of the paper.

$$P\sqrt{L^2 - d^2} = mg\frac{d}{3} + Mg\frac{d}{2}. \tag{21}$$

Thus we have three equations for the three unknowns R, φ, and P and can solve these equations for whichever quantity we want.

PROBLEMS

1. At a place where $g = 9.8$ m./sec.^{2}, a body is thrown vertically downward with a velocity of 1 m./sec. How far will it have gone in 5 sec.?

2. At a place where $g = 9.8$ m./sec.^{2}, a body is thrown vertically upward with a velocity of 30 m./sec. How far from the ground will it be in 5 sec.?

3. A ball is thrown vertically upward with a velocity of 15 m./sec. One second later, a second ball is thrown vertically upward with the same

velocity. How far from the ground and after how long a time will the two balls meet?

4. In a head-on automobile collision on a level road, the glass of the windshield from one car was found 15 m. beyond the point of collision. Assume that the glass started as a projectile with the velocity of the car in a horizontal direction at the time of collision and that the windshield is 1.5 m. above the road, and calculate the velocity of the car at the time of collision.

5. Two stones are thrown from the top of the Empire State Building 380 m. from the street at the same instant. One is thrown horizontally with a speed of 30 m./sec. and the second one upward at an angle of 45° with a velocity of 30 m./sec. What is the difference in time and position at which they reach the ground?

6. An airship dropped a wrench from directly above the laboratory. If the airship was traveling at 80 mi./hr. and the wrench landed 1,200 ft. from the laboratory, what was the height of the airship?

7. An anti-aircraft gun fires at a moving target at the moment that the target is overhead. If the height of the target is h, its velocity w, and the muzzle velocity of the projectile V, at what angle must the gun be pointed to score a hit on the upward flight of the projectile?

8. Suppose a ski jumper takes off horizontally with a speed of 16 m./sec. If the landing slope starting from the take-off makes an angle of 45° with the horizontal, how far down the slope from the take-off will he land, and how long will he be in flight?

9. A certain roof makes an angle of $\theta°$ with the horizontal. If a mass of M kg. of ice becomes dislodged at a point h m. from the eaves and if the height of the eaves above the ground is H m., how far from the house will the ice strike the ground? Neglect friction.

10. It takes a juggler $\frac{1}{2}$ sec. to catch a ball, toss it up again, and be ready for the next. How high must he throw the balls to keep four going at once?

11. What would be the range of the gun of illustrative Prob. 1, page 93, if it were located on a level plane on the surface of the moon?

12. The world's record of 26 ft. $8\frac{1}{4}$ in. for the broad jump was set by Jesse Owens at Ann Arbor on May 25, 1935. How much error would be made in this record by having the level of the dirt in the pit 1 in. below the level of the take-off? Assume an angle of take off of 30°.

13. The world's 16-lb. shot record of 57 ft. 1 in. was set by Jack Torrance at Oslo on Aug. 4, 1934. If the value of g at Oslo is 9.819 m./sec.^{2}, what would this record have been at Los Angeles where g = 9.796 m./sec.^{2}? Assume the shot put at an angle of 45° with the horizontal 5 ft. from the ground.

14. The world's javelin record of 251 ft. 6 in. was set by Matti Jarvinen at Turin, Italy, on Sept. 7, 1934. At Turin, the value of g is 9.805 m./sec.^{2} What would have been the world's record if the javelin had been thrown at Berlin where g = 9.813 m./sec.^{2}? Assume the javelin thrown at an angle of 45° with the horizontal 5 ft. from the ground.

15. Find the momentum of a 2-kg. mass which has fallen freely from rest for 5 sec.

16. A man weighing 78 kg. is accelerated upward by an elevator 3 m./sec.^{2} Find the magnitude and direction of the force that the floor of the elevator exerts on him.

17. A ball of mass M and velocity V strikes a wall perpendicularly and rebounds with undiminished speed. Find the average force exerted on the wall during the time of collision t.

18. A rubber ball having a mass of 60 g. hits a wall with a speed of 13 m./sec. Assume that it bounces off with a speed 0.8 of its original speed. Find the average force exerted on the ball by the wall if the time of contact is 0.01 sec., if it is 0.0001 sec.

19. Suppose that a stream of mercury molecules is emerging from a small furnace in an evacuated vessel. They hit a light vane suspended in the stream. If the mass of each molecule is 3.2×10^{-25} kg., its velocity 3×10^4 m./sec., and 10^{17} molecules strike the vane per second and condense on it, how much momentum is given to the vane per second? What is the average force acting on the vane? What would the force be if the molecules bounced back off the vane with a speed equal to their initial speed?

20. If the brakes of an automobile having a mass of 1,500 kg. can exert a force of 9,800 newtons, in how short a distance can the car be stopped if it is traveling 90 km./sec.?

21. A 100-kg. block of iron rests on a smooth horizontal table. A string attached to the iron runs over a smooth pulley at the edge of the table and carries a 15-kg. weight which hangs vertically. If the iron is initially at rest 10 m. from the pulley, how long will it be before the iron reaches the pulley and how fast will it be moving? Neglect friction and the inertia of the pulley.

22. Two weights of 6 kg. each are connected by a cord and suspended from a pulley. The mass of the cord and the mass and friction of the pulley may be neglected. If a weight of 0.1 kg. is added to one side, find the downward acceleration of that side. Also find the tension in the string.

23. A car starts from rest on a track inclined at an angle of 30° to the horizontal. If the mass of the car is 500 g., what are the components of the force of gravity acting on the car parallel and perpendicular to the track (give answer in newtons)? What is the acceleration of the car, and how far does it go down the track in 2 sec.? Neglect friction in this problem.

24. If the table of Prob. 21 is tilted at an angle of 20° to the horizontal and the string goes over its upper edge, what are the acceleration and the tension in the string?

25. If a train leaving New York reaches its constant running speed of 70 mi./hr. in 4 min. and has 10 cars each of mass 80 tons, what must be the tractive force exerted by the engine?

26. What is the force of gravitational attraction between two spheres of lead of radii 2 and 20 cm., respectively, when they are just in contact? Between two liners each of 100,000 tons mass and with their centers of mass 500 ft. apart?

27. How far above sea level must a body be in order that its weight will be reduced to one-fourth its weight on the earth's surface?

28. The mass of the moon is $\frac{1}{80}$ that of the earth and its radius is 1,770 km., whereas the earth's radius is 6,357 km. Find the acceleration of a freely falling body on the surface of the moon.

29. On the moon, the acceleration of a freely falling body is about one-sixth its value on the earth. Find the time it would take an object to fall from the top to the bottom of a 50-m. cliff on the moon.

30. Jupiter has a mass about three hundred times that of the earth and a radius of about eleven times that of the earth. If a man weighs 80 kg. on the earth, how much would he weigh on Jupiter?

31. If the mass of the earth is 5.95×10^{24} kg., the mass of the moon = 7.3×10^{22} kg., the radius of the moon's orbit 3.85×10^{8} m., and the numerical value of the gravitational constant is 6.66×10^{-11} newton-m.^{2}/kg.^{2}, find the force between the earth and the moon.

32. Find the acceleration of a falling meteor when it is at a distance from the earth's surface equal to twice the earth's radius.

33. If the earth attracts a mass of 1 kg. with a force of 9.80 newtons, what is the force of attraction between two masses each of 1 kg. placed 1 m. apart? The mass of the earth is 5.95×10^{24} kg., and its radius 6.36×10^{6} m.

34. If the density of lead is 11,350 kg./cu. m., find the mass of one of the weights and one of the balls used in Cavendish's experiments.

35. If in Cavendish's experiment the force between the weights and the lead balls is $10^{-6} \times 1/48$ the weight of a ball when the distance between weights and balls is 18 cm., find the mass of the earth and its density.

36. From the mass of the earth and its radius, find the value of the universal constant of gravitation G.

37. If a flea can jump 1 m. high on the earth, how high could he jump on the moon? How long would he be off the ground in each case?

38. Knowing that an acceleration toward the earth was necessary to keep the moon in her circular path and that the radius of the moon's orbit was sixty times the radius of the earth, Newton calculated that the moon ought to fall toward the earth with a certain acceleration. Find this acceleration.

39. Forces of 4- and 5-kg. weight make an angle of 60° with each other. Find their resultant.

40. If a particle is in equilibrium under the action of three forces of 7-, 10-, and 13-kg. weight, find the angles between them.

41. A man is rolling beer barrels up an inclined plane onto a truck. How much harder does he have to push if he pushes horizontally than if he pushes parallel to the plane? How much more force would be required to lift the barrel straight up than to push parallel to the plane?

42. Two groups of six men each are engaged in a tug of war, each man pulling 1,000 newtons. Find the tension in the rope.

43. A bird alights on the center of a telephone wire. The bird weighs 200 g., the telegraph wire is 30 m. long and is depressed 5 cm. from the horizontal. Find the tension in the wire produced by the bird. The weight of the wire may be neglected.

44. Two tugs A and B are attached to the front of a barge by separate ropes. If tug A is pulling twice as hard as the other and the angle its rope

makes with the direction of motion of the boat is 15°, in what direction is the other tug pulling?

45. If the force necessary to move the barge of Prob. 44 is 10,000 newtons, what are the pulls exerted by the tugs?

46. Explain how it is possible for a sailboat to move into the wind.

47. A weight hung on the middle of a pole 4 m. long is being carried by three men. One man carries one end of the pole. The other two carry each end of a cross bar on the center of which the pole rests. How far from the weight must the cross bar be placed in order that all the men carry the same weight?

48. Find the tension T in the rope of the crane in Fig. 40. Assume compression only in the boom, and neglect its weight. If the boom is uniform

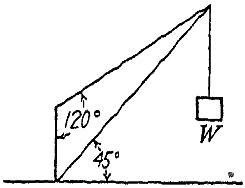


FIG. 40.—Crane. Prob. 48.

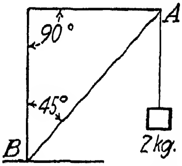


FIG. 41.—Derrick. Prob. 50.

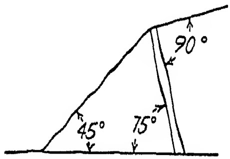


FIG. 42.—Stake for a circus tent. Prob. 51.

and has a weight w, what will the tension in the rope be? The rope is securely fastened at the top of the boom.

49. A 10-m. ladder rests against a smooth vertical wall. The distance of its base from the wall is 1 m. If the weight of the ladder is 20 kg. and if it is uniformly distributed, calculate the force that it exerts against the wall.

50. If the weight of the beam AB, Fig. 41, is negligible, find the force of compression on it. The rope is securely fastened at the top of the beam AB.

51. A pull of 500-kg. weight by a circus tent is held by a rope from a stake to which it is perpendicular, Fig. 42. The stake makes an angle of 75° with the ground, and the stay holding it makes an angle of 45° with the ground. If the stake is free to rotate about its lower end, find the tension in the stay.

CHAPTER V

WORK AND ENERGY

1. The things that men need to do in order to provide for their material existence involve the doing of work. What is work? Work may be done in a variety of ways, but one of the commonest ways is that of lifting weights. As we derived the notion of force from the notion of muscular effort, so also we can obtain a clear picture of the physicist's concepts of "work" and of "energy" by consideration of the work done by human beings in simple mechanical tasks.

2. If a weight of a certain number of pounds is to be lifted from one level to another, we feel intuitively that the quantity of work involved will be proportional to the number of pounds to

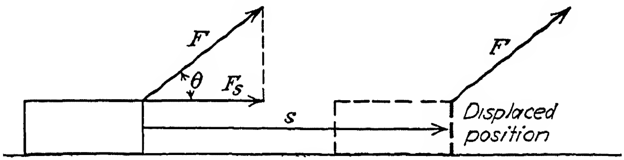


FIG. 43.—The work done by a force F, in causing a horizontal displacement s, is $s\,F\cos\theta$, where θ is the angle that F makes with s.

be lifted and to the distance through which it is lifted. That is, to lift 200 lb. through 5 ft. will be naturally regarded as taking 200×5 or 1,000 times as much work as if we raised only 1 lb. through 1 ft. That is the definition that we shall adopt as our measure of work.

The work done in moving a body is defined as equal to the product of the component of force in the direction of the displacement and the amount of the displacement.

Work may be measured therefore in newton-meters or foot-poundals if absolute units of force are used. The newton-meter is the amount of work done when a force of 1 newton moves a body through a distance of 1 m. It is also called the joule.*

* In the c.g.s. system, the unit of work is the dyne-centimeter and it is called the erg. One joule equals 10^7 ergs.

If gravitational units of force are used, work is naturally measured in foot-pounds, one foot-pound being the amount of work necessary to lift a weight of one pound through a vertical height of one foot. Similarly in the metric system, one kilogram-meter is the work necessary to lift one kilogram a vertical height of one meter.

3. The capability of doing work we shall call *energy*. In order to be able to lift a weight, one must possess energy. It is natural to measure quantity of energy, *i.e.*, quantity of the capability of doing work, in terms of the quantity of work one is capable of doing. Thus if a certain agency is capable of doing 50 newton-m. of work, we say that it must possess at least 50 joules of energy.

4. It is characteristic of all sources of energy of which we know that they are used up by doing work. That is, if we have some source of energy which is capable of lifting weights, it will have only a limited capability of doing this. After having lifted weights to the extent of a certain number of newton-meters, it loses its capability for further doing of work and is exhausted. Any apparatus therefore has at any time a finite store of energy, just as it has a finite amount of matter in it. If the apparatus is made to do work to the extent of a certain number of newton-meters, then its store of energy is diminished by just this amount. No energy is actually lost, but it is converted into less available forms.

5. Much human effort has gone into the attempt to discover a so-called *perpetual-motion* machine. By this is meant a machine that is capable of doing an unlimited amount of work without diminution of its energy or capability of doing further work. The name "perpetual motion" is something of a misnomer. We have plenty of examples of perpetual motion in nature in the sense of a motion that keeps on uniformly without any sign of slowing down: the rotation of the earth on its axis is a case in point. What is really desired by perpetual-motion inventors is a machine that not only runs continuously when left alone, but also may be made to drive other machinery, thereby doing work, without any diminution in its ability to keep on going. Obviously, such a machine would be very useful. Despite great efforts to build such a machine, none has ever been devised so that the impossibility of constructing a perpetual-

motion machine may be regarded as a well-established law of physics.*

6. Let us make the discussion a little more concrete by thinking of a definite machine for lifting weights. We may take an old alarm clock and arrange to have the train of gears from the clock spring turn an axle with a small windlass on it on which a thread may wind up. If we hang a small weight on the end of the thread, the clock mechanism will cause the windlass to wind up the thread, thereby lifting the weight. In other words, the clock mechanism may be made to do a definite amount of work. This work is measured in newton-meters, the product of the weight in newtons by the distance the weight is raised in meters.

7. But everyone knows that the clockwork derived this capability for doing work by the fact that someone has wound up the spring. In other words, the clockwork is not a *source* of energy but merely a device for *storing* energy. For it takes work to wind up the spring, and this has to be supplied by the muscular effort of the person who twists the winding key. The work that he does in winding the spring is equal to the product of the force that his fingers exert against the winding key multiplied by the distance through which they move while winding up the clockwork.

8. The lifting of the weight by the clockwork is done at the expense of an unwinding of the spring, *i.e.*, a diminution in the capability of doing further work possessed by the clockwork, in other words, a diminution of the energy possessed by the clockwork. So the clockwork loses energy to the extent that it does work in lifting the weight.

9. What becomes of the energy that the clockwork loses in lifting the weight? It went to do the work necessary to lift the weight. But is it completely lost or used up? No, it is merely stored in another form. Evidently if we have a raised weight, it has a capability for doing work to drive a mechanism or to lift another weight. In fact, a very common type of "grandfather's clock" uses this form of storing energy rather than a spring. In such clocks, the driving energy is supplied not by winding up a spring but by lifting a weight once a day or once a week. The weight then slowly descends and in its

* For an interesting historical account of the attempts at building perpetual-motion machines see Dirck, "Perpetual Motion."

descent supplies the driving energy needed to keep the clock running.

10. We are thus already familiar with two forms in which energy may be stored.

1. *Elastic Energy.* An elastic body, for example a clock spring, has a certain "natural" shape which it assumes when no forces act on it. In order to distort it away from this shape, forces are needed and as these forces move in the act of distorting the elastic body they do work on the body. The distorted body thereby acquires the capability of doing work. Because of its tendency to resume its original shape, it can exert forces which are able to lift weights. Its capability of doing work is thus a form of energy which comes from the fact that it has been distorted. Such energy is usually called elastic energy or energy of deformation.

2. *Gravitational Potential Energy.* This is the name that physicists give to the energy which a body possesses in virtue of being in an elevated position so that it may do work by falling. When a weight is lifted by our clockwork mechanism just described, we say that the energy has been transformed from the elastic energy which was in the wound-up spring into the gravitational potential energy which the weight has in virtue of its elevated position.

11. It is interesting to get a general relation for the amount of gravitational potential energy possessed by a body. In Par. 6, it has already been pointed out that the work done in lifting a body is the product of the weight and the distance it is raised, *i.e.*, $W = mgh$ if h is the height through which the mass m is raised and W is the work done. This amount of energy has been stored in the body and can be recovered if the body is allowed to fall. We say therefore that the body has a potential energy of mgh. Obviously, the amount of this energy depends on h and therefore on the horizontal plane from which h is measured. Thus gravitational potential energy is always expressed in terms of some reference plane, usually the earth's surface.

12. There are many other forms in which energy, or capability of doing work, may appear in nature such as heat energy, chemical energy, electrical energy, etc. Physics is largely concerned with these forms and the manner in which they are transformed into one another. Let us now go on to consider another form in which energy may appear in nature.

13. Suppose that we have a body of mass 10 kg. moving along with a velocity of 5 m./sec. The quantity of motion or momentum that it has is then 50 kg-m./sec., as we saw in the last chapter. We know that we have to apply a force to a body to change the quantity of its motion. In particular, if we wish to stop the body, we shall have to pull back on it with a constant force F in newtons for a time t in seconds such that $Ft = -mv$ where mv is the quantity of motion it possesses. For we saw in Chap. IV, Par. 6, that

$$F_{av} = \frac{M_t - M_0}{t - t_0} \tag{1}$$

so that if the momentum changes from $M_0 = mv$ to $M_t = 0$ in the time interval $t - t_0 = t$ sec., $Ft = -mv$. The minus sign shows that the force and the original momentum are in opposite directions. We might pull back on the body by attaching a spring to it and holding fast to one end. As the body moves on, it will have to stretch the spring until it is brought to rest with the spring stretched. Now we know that a stretched spring contains elastic energy. In other words, the body had to do work on the spring in order to stretch it. The body possesses this ability to do work simply by virtue of its motion.

14. Thus we see that a body in motion has a capacity for doing work which it possesses simply by virtue of its being in motion. This is a new form of energy to add to our list. Scientists call it *kinetic energy*. The question now comes up: How much kinetic energy does a body of a certain mass have when it is moving with a given velocity? We might devise experiments to measure this energy in its dependence on the mass and the velocity of the body. For example, we could make the moving body stretch a spring until it was just brought to rest, and then we could measure the ability of the stretched spring to lift weights and so get a measure of the energy. Or it might be possible to arrange matters so that the body in being brought to rest lifted a weight directly, and then the amount of the kinetic energy would be its ability to do work in lifting this weight.

15. But we can also find how the kinetic energy depends on the mass and velocity by a logical argument from the laws of motion which we learned in the preceding chapter. Returning to our 10-kg. body going 5 m./sec., we know that a force of 20 newtons will bring it to rest in 2.5 sec. During this time, its speed will

diminish uniformly from 5 m./sec. down to 0 m./sec., and so it will travel with an average speed of 2.5 m./sec. Therefore the distance that it moves while being brought to rest is 2.5×2.5 or 6.25 m. In other words, the body was able to keep going a distance of 6.25 m. in spite of the retarding action of a back pull of 20 newtons acting on it. It managed to pull against a force of 20 newtons for a distance of 6.25 m. The amount of work that it did was therefore 20×6.25 newton-m. As it is now at rest, it has no motion and so can have no kinetic energy. The kinetic energy, which is its capacity for doing work, was therefore 125 newton-m. or joules.

16. You may feel that this result is very special and depends on the fact that the body was being brought to rest by a force of a particular value. But it is easy to see that that is not true. For if we make the body work against a force twice as great, it will be brought to rest in half the time. Its average speed will be the same, but in half the time it will go only half as far. When figuring up the work done, the result comes out the same, for we now have twice the force working through half the distance, which leaves the product unaltered.

17. Let us now see how the result works out for a body of quite general mass m kg. moving with a velocity v m./sec. If we apply a steady retarding force F newtons, we know from the last chapter that it will be brought to rest in t sec., where t is given by

$$t = \frac{mv}{F}. \tag{2}$$

Its average velocity will be the average of v and 0 which is $v/2$ m./sec.; so the distance it goes in the time t is $vt/2$ which becomes $mv^2/2F$ if the result of Eq. (2) is used. The work that the body does is the product of the distance gone by the amount of the force it works against. This is evidently $F \times mv^2/2F$ or just $mv^2/2$. The fact that F cancels out in this argument is just the mathematical expression of our argument of the preceding paragraph. The greater the force, the more quickly the body is brought to rest, and hence the less distance the body moves in being brought to rest; but nevertheless, the product of force times distance gone which is the work done comes out quite the same.

18. *We have thus learned from what we already know about the laws of motion that the kinetic energy of a body of mass m and velocity v is equal to $mv^2/2$.* If the mass is in kilograms and the velocity is in meters per second, we know that F must be measured in newtons in the equation $Ft = mv$. Hence, in this system of units, we are led to the answer $mv^2/2$ expressed in newton-meters or joules. If the mass is in pounds and the velocity is in feet per second, the energy comes out measured in foot-poundals, which is the amount of work done when a force of one poundal moves a body through a distance of 1 ft.

The preceding discussion has therefore made us familiar with a third form of energy which we call kinetic energy.

19. Let us recapitulate the kinds of units of energy that we have found convenient to introduce.

Newton-meter: Amount of energy expended when the point of application of a force of one newton moves through a distance of one meter parallel to the direction of the force. Also called *joule.*

Kilogram-meter: Amount of energy that is capable of lifting one kilogram vertically through one meter.

Foot-poundal: Amount of energy expended when the point of application of a force of one poundal moves over a distance of one foot parallel to the direction of the force.

Foot-pound: Amount of energy that is capable of lifting one pound vertically through one foot.

20. Of these units, the foot-poundal and the newton-meter, or joule, are more fundamental than the foot-pound or the kilogram-meter. In the preceding chapter, we have seen that the force which is a newton is a definite amount of force, the same at all places. But the force that is the weight of one kilogram is a variable thing, depending on the location on the earth since it not only depends on the amount of matter, which is being acted on, but also on its relation to the attracting body, the earth, which does the pulling. Hence, to be quite precise, when we speak of one kilogram-meter, we should be careful to say at what place on the earth the one-kilogram mass was lifted through a height of one meter, for this will be different at different places. No such difficulty attached to the use of the newton-meter, for a newton is defined in a way that does not make it depend on a variable factor like the earth's attraction. The same remarks apply to the corresponding English units. The foot-pound is an

amount of energy that varies from place to place according to the strength of the earth's attraction for one pound mass at that place. The foot-poundal, however, is defined in a way that is independent of the earth's attraction and so is a more fundamental unit.

21. We have already seen in our introductory note on units that the *joule* is about the size of the amounts of energy one deals with in small mechanical operations. It is easily seen to be approximately the work necessary to raise a weight of 1 kg. (2.2 lb.) through a distance of 10 cm. (4 in.). Another unit in common use is the *kilowatt-hour* which is much larger. It is the unit of energy in which electric-power bills are always expressed; hence one often sees it on the front page of the newspapers in articles where regulation of public utilities is being discussed. The *watt* is a unit of *power*, which is the *rate of doing work*. A machine is delivering power of one watt when it is doing work at the rate of one joule per second. As in "kilogram" or "kilometer," the prefix kilo- means one thousand; a kilowatt is a rate of doing work which amounts to 1,000 joules per second. A kilowatt-hour is the amount of work done by a machine in one hour when it is working at the rate of one kilowatt. As there are 3,600 seconds in an hour, it is therefore 3.6 million joules. A typical small household uses about 50 kw.-hr. of electrical energy per month for lighting, heating, and small domestic mechanical contrivances.

22. In explaining the origin of the energy unit *kilowatt-hour*, we were led to touch briefly on the concept of *power* which is closely related to that of work or *energy*. In practical affairs, it is not enough that work be done: we usually are interested in having it done in a certain time. Therefore we wish to rate our engines by the time rate at which they are able to supply energy. There are various units of power in use. One of these is the watt (and the thousand-fold multiple, the kilowatt) already mentioned. Another is the "horsepower." This is a very old unit, going back to the days before the modern machine age in which horses were used as a means of getting mechanical work, and was standardized as the rate of doing work of which a good average work horse is capable. Of course, horses are variable, and so one does not keep a standard horse in the Bureau of Standards as a means of giving precise meaning to one horsepower. What was

done long ago was to agree arbitrarily on a value that corresponds roughly to what a horse can do. This value is 550 ft.-lb./sec. In terms of the modern power unit, this is about $\frac{3}{4}$ kw., more precisely 746 watts. Now that horses are used very little as "prime movers" in technical work, the relation to horses of 1 hp. has little more than historical interest.

23. What is perhaps of more interest is a feeling for the rate at which human beings do work in various processes. Such measurements have been made by research workers in physiology, for they are of great importance in trying to get at the precise nature of the way in which the chemical energy stored in food is transformed into ability to do work in muscular processes. A few results of such measurements on man made by physiologists are given in the following table.

Type of activity	Duration	Horsepower	Watts
Mountain climbing:			
Moderate	Many hours	0.11	82
Severe	1 to 2 hr.	0.19	142
Very severe	4 min.	0.43	320
Climbing treadmill	30 sec.	0.65	485
Running upstairs	30 sec.	0.93	700

24. While we are on this subject of muscular work done by man or his rate of doing muscular work, let us consider one point that often gives students trouble. Physicists define work as the product of force multiplied by the distance through which it moves. Therefore in the physicists' sense of the word work, no work is done in merely holding up a weight without moving it. Yet to hold a weight of 1 lb. out at arm's length horizontally for even a few minutes is a very fatiguing task. It is "work" in the colloquial sense of the word and yet, since the distance moved appears to be zero, it would seem that our muscles have done no work in the process. The answer to this difficulty is that the mechanism whereby our muscles exert a static, nonmoving force is not so simple as this description implies. In physiological studies, it is shown that muscle activity, even for the production of forces which do not bring about motion of the arms or legs as a whole, is the average effect of many small twitching contractions of individual muscle cells which are brought about by the con-

sumption of the energy stored chemically in the material of the tissue. But the subject of the exact way in which the energy value of food is converted into mechanical energy by muscular action is one of the most complicated problems in the science of physiology, and so we must not attempt any more than this hint at the way in which the paradox is explained.

Friction and the Conservation of Energy

25. In Par. 5 of this chapter, we stated, perhaps too glibly, that we had plenty of examples in nature of perpetual motion but that a perpetual source of energy had never been discovered. The reason that the first part of the statement was a trifle optimistic is that it is very difficult to eliminate all forces acting on a moving system so that motion usually involves the doing of work. The forces that are hard to eliminate are the forces of friction, the forces tending to prevent the sliding of one surface over another. In the broadest sense, this includes not only the surfaces of contact between solids as in sliding or rolling mechanical friction, but also the surfaces of contact between solids and fluids and between adjacent layers of fluids themselves such as are involved in motion through water or air or in the flow of water or air. In the last analysis, these frictional forces are an expression of the forces of attraction between the molecules of the substances involved, the same type of force that we discussed in connection with surface tension and elasticity. The energy used in overcoming them appears in the increased motion of the molecules, *i.e.*, in the generation of heat, familiarly experienced as the result of friction. Practically it is impossible to deal with frictional forces in detail as intermolecular forces. Instead, the results of experiment are incorporated in certain empirical laws and numerical coefficients such as are discussed in the next chapter.

26. In the first chapter, the principle of the conservation of mass was discussed and the analogous principle of the conservation of energy foreshadowed. In order to test the principle that energy can be neither created nor destroyed in any physical process, we must be able to measure equivalent amounts of energy in different forms. If some energy has gone into the kinetic energy of molecules, proper allowance must be made for it by measuring the change in the temperature of the system. We are not yet in a position to do this since we have not yet con-

sidered the subject of heat. But if the friction in a given mechanical experiment can be made sufficiently small, the energy converted into heat can be neglected and all the energy will remain mechanical, *i.e.*, potential, elastic, or kinetic. This principle is particularly useful in studying questions having to do with falling bodies. In such problems, the sum of the potential and the kinetic energy is constant. Thus, if a body starts from rest and falls through a height h, its energy originally is all potential and equal to mgh. Therefore at the end of its fall, its kinetic energy is $\frac{1}{2}mv^2 = mgh$, according to the principle of conservation of energy. A bouncing ball is continually changing its energy from potential to kinetic, to elastic, to potential, and so on, until friction stops it. The kinetic energy of a projectile is $\frac{1}{2}mv^2$ when it leaves the gun if v is the muzzle velocity. At the top of the trajectory, this has been partially converted to potential energy but goes back to kinetic energy as the projectile falls. Often problems can be solved much more simply by the application of the principle of conservation of energy than by studying the forces and accelerations in detail. Though at present we are dealing only with mechanical energy, we can just as well state the Principle of Conservation of Energy at this point in general form. It is:

The total amount of energy in any system is unchanged by physical or chemical processes within that system.

Energy may be transformed from one kind to another, but it can be neither created nor destroyed. This is one of the most fundamental principles of physics, perhaps the most fundamental. It is constantly used in the interpretation of general phenomena and in the solution of specific problems. One such application is in the treatment of collisions which now follows. Other examples will be found in the illustrative problems and the problem set at the end of the chapter.

Collisions

27. Although we have already discussed all the principles needed for the study of collisions between moving bodies, such phenomena are of sufficient importance to warrant special treatment.

28. We have already considered a bouncing ball as an example of the change of energy from one form to another. That was an

example of the collision of one body with another of very much larger mass, practically infinite. Now consider a collision between two bodies of comparable mass such as two billiard balls. Suppose they are making a head-on collision. Before they come in contact, each ball has kinetic energy. During contact, each ball exerts a force on the other. The effect of this force on each ball as a whole is to reduce its velocity, perhaps even reversing the direction of motion. The effect on the molecules near the region of contact is a displacement from their normal positions, an elastic distortion. If the balls are elastic as billiard balls are, there will be an elastic force which will tend to restore the molecules to their normal position. This elastic restoring force will push the balls apart, acting until they are no longer in contact and giving them velocity so that they bounce away from each other. In terms of energy, some of the original kinetic energy of the balls goes into elastic energy of distortion and then is reconverted into kinetic energy as the balls recover their original shape. If the balls are "perfectly elastic," all the elastic energy has been reconverted into kinetic energy when the collision is over. In fact, this is the best definition we can give of a "perfectly elastic" collision. In practice, even the most elastic of materials will change some of the original kinetic energy into irrecoverable heat energy in the process of elastic distortion and recovery.

29. Clearly, then, what happens at a collision depends on the elastic properties of the colliding bodies. In practice, we encounter all degrees of elasticity in collisions and can express them in terms of a so-called coefficient of restitution. In this book we shall consider only the two limiting cases.

30. In one case, which we call a *perfectly elastic collision*, we assume that the colliding bodies return to their original shapes and that the heating produced is negligibly small. In practice, this situation is closely approximated in the collision of two ivory balls, slightly less so in the collision of two hard steel balls. Since no energy goes into permanent distortion or heating, the total kinetic energy of the bodies after the collision must be the same as before.

31. In the other case, which we call a *perfectly inelastic collision*, we assume that the elastic restoring forces are negligibly small, so that there is nothing tending to push the bodies apart after the

collision. They will therefore move along together with a common velocity. These conditions are closely approximated by lumps of putty or bags of sand. For such materials there will be permanent distortion of the colliding bodies and energy lost in heat so that the total kinetic energy of the bodies after the collision will be less than before the collision.

32. For any collision, elastic or inelastic, the two colliding bodies form a moving system on which no external force is acting. Therefore (see Chap. IV, Par. 24) their *total resultant momentum remains constant, i.e.,* the resultant momentum after the collision is the same as before the collision.

33. We are now in a position to consider the two cases mathematically. Suppose we have two bodies of masses m_a and m_b.

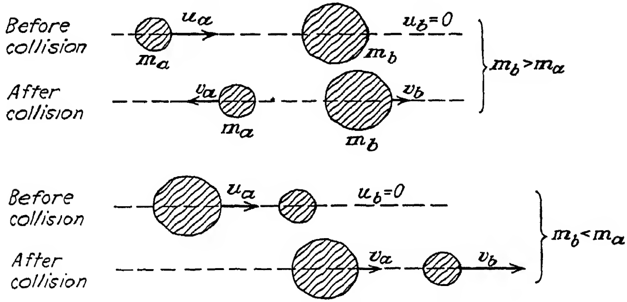


FIG. 44.—Perfectly elastic collisions.

Let their velocities before collision be u_a and u_b along the same line and after collision be v_a and v_b. Let the positive direction be to the right and the negative to the left as usual. If the collision reverses the direction of motion of one of the bodies, the velocity will change sign.

1. *Perfectly Elastic Collisions.* From the conservation of momentum we have

$$m_a u_a + m_b u_b = m_a v_a + m_b v_b. \tag{3}$$

From the definition that an elastic collision is one where no kinetic energy is lost,

$$\tfrac{1}{2}m_a u_a^2 + \tfrac{1}{2}m_b u_b^2 = \tfrac{1}{2}m_a v_a^2 + \tfrac{1}{2}m_b v_b^2. \tag{4}$$

These equations relate the masses, initial velocities, and final velocities. If all but two of these quantities are given, the other

two can be found. We shall carry out the solution for the particular condition where one of the bodies, say m_b, is initially at rest, *i.e.*, $u_b = 0$. Therefore,

$$m_a u_a = m_a v_a + m_b v_b \tag{5}$$

and

$$m_a u_a^2 = m_a v_a^2 + m_b v_b^2. \tag{6}$$

Suppose we eliminate v_b and solve for v_a, we get

$$v_a = \frac{m_a - m_b}{m_a + m_b} u_a. \tag{7}$$

Evidently this can be either plus or minus depending on whether m_a is greater or less than m_b. Thus a body moving before collision

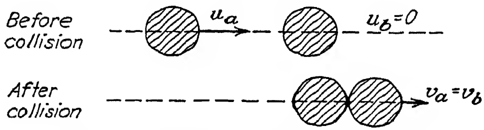


FIG. 45.—Perfectly inelastic collisions.

may continue its forward motion or bounce back, reversing the sign of its velocity. Or, if we solve for v_b, we get

$$v_b = \frac{2m_a}{m_a + m_b} u_a \tag{8}$$

which must always be positive.

2. *Perfectly Inelastic Collisions.* As before, the momentum remains constant so that

$$m_a u_a + m_b u_b = m_a v_a + m_b v_b,$$

but now the kinetic energy does not remain constant so that we have no relation between the kinetic energy before and after collision. We have, however, the fact that the two bodies move together after a perfectly inelastic collision (by definition). Therefore,

$$v_a = v_b.$$

Again assuming that $u_b = 0$, we have

$$m_a u_a = m_a v_a + m_b v_b$$

or since $v_a = v_b$

$$m_a u_a = (m_a + m_b) v_b \tag{9}$$

or

$$v_b = \frac{m_a}{m_a + m_b} u_a \tag{10}$$

which is always positive and less than u_a.

SUMMARY

The work done in moving a body is the product of the displacement and the component of the force parallel to the displacement. Energy is the capacity to do work. The unit of work or energy in the m.k.s. system is 1 newton-meter and is called a joule. Energy is never created or destroyed, but it changes from one form to another. It may exist as elastic energy, gravitational potential energy, or in many other forms. Kinetic energy is the energy that bodies have by reason of their motion; it is equal to $mv^2/2$. Power is defined as rate of doing work. The unit of power in the m.k.s. system is one joule per second and is called the watt. Human beings in various activities perform work at the rate of between 50 and 1,000 watts. In all practical machines, some energy is required to overcome friction and is converted into heat energy. Though such energy cannot be recovered as useful work, it is not destroyed. In fact, one of the most fundamental principles of physics is the principle of conservation of energy according to which:

The total amount of energy in any system is unchanged by physical or chemical processes within that system. Energy may be transformed from one form to another, but it can be neither created nor destroyed.

The motion of two bodies after collision can be computed in terms of their masses and their motion before collision by applying the principles of conservation of energy and conservation of momentum. Two extreme cases are considered: perfectly elastic collisions in which both momentum and kinetic energy are conserved, and perfectly inelastic collisions in which momentum but not kinetic energy is conserved and in which the two bodies move together after collision. The equations for both cases are derived.

ILLUSTRATIVE PROBLEMS

1. A man pulls a sled by a rope making an angle of 45° with the horizontal. If he exerts a force of 40-kg. weight and pulls the sled 1 km., how much work does he do?

The part of a force that is effective in doing work is the component of the force in the direction of the motion. The component of the force at right angles to the direction of the motion does no work. Therefore we must find the component of the 40-kg. weight in the direction of motion of the sled, *i.e.*, in the horizontal direction. This component is 40 kg. $\times \cos 45° = 28.3$ kg. wt. The work is then the product of this force and the distance.

$$W = 28.3 \text{ kg.} \times 1{,}000 \text{ m.} = 28{,}300 \text{ kg.-m.}$$

In m.k.s. units the force is $9.8 \times 28.3 = 277$ newtons and the work is 277,000 joules.

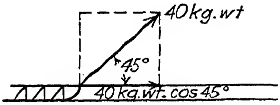


FIG. 46.—Forces acting on a sled for the computation of work.

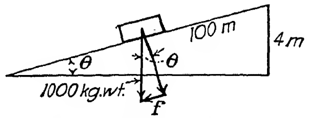


FIG. 47.—An automobile climbing a hill. f is the component of the force of gravity opposing the motion.

2. An automobile weighing 1,000 kg. and traveling 80 km./hr. on the level maintains the same speed after starting up a grade rising 4 m. per 100 m. of road. Find the extra power required.

Power is the rate of doing work. The work is the product of the force in the direction of motion times the distance moved. Therefore the power is $F \times d/t = Fv$ since d/t represents the average velocity at which the distance d is traversed. The force must be constant during the time interval t or its average value used. The additional power on the upgrade is that required to overcome the component of the force of gravity which acts parallel to the road down the hill. This force is seen from Fig. 47 to be $f = 1{,}000 \text{ kg.} \times 9.18 \text{ newton/kg.} \ldots \sin\theta$ or since $\sin\theta = 4 \text{ m.}/100 \text{ m.} = .04$, $f = 392$ newtons. The power required to maintain a speed of 80 km./hr. against this force is

$$\begin{aligned} P &= 80 \text{ km./hr.} \times 392 \text{ newtons} \\ &= 80/3600 \text{ km./sec.} \times 392 \text{ newtons} \\ &= 8.71 \text{ newton-km./sec.} \\ &= 8.71 \times 10^3 \text{ newton-m./sec.} \\ &= 8.71 \times 10^3 \text{ joules/sec.} \\ &= 8.71 \times 10^3 \text{ watts} \\ &= 8.71 \text{ kw.} \end{aligned}$$

3. A motor truck of 4,000-kg. mass is starting to tow another car of 2,000-kg. mass. If the towing truck attains a speed of 15 km./hr. before the tow rope becomes taut, how fast are the two cars moving just after the rope becomes taut?

As the rope is assumed not to stretch, this is a case of inelastic collision. The momentum of the towing truck before collision is therefore equal to the sum of the momenta of the two cars after collision. Their velocity after collision is the same. Thus:

$$m_a v_a = (m_a + m_b)v_b$$
$$15 \text{ km./hr.} \times 4{,}000 \text{ kg.} = (4{,}000 \text{ kg.} + 2{,}000 \text{ kg.})v_b$$
$$v_b = \frac{15 \text{ km./hr.} \times 4{,}000 \text{ kg.}}{6{,}000 \text{ kg.}} = 10 \text{ km./hr.}$$

PROBLEMS

1. A mass of 150 lb. is lifted 40 ft. Calculate the work done in foot-pounds, foot-poundals, kilogram-meters, and joules.

2. A mass of 250 kg. rests on a smooth horizontal surface. A force of 40 newtons is applied to it for 12 sec. Find

a. The acceleration.

b. The final velocity.

c. The distance away from the starting point at the end of the 12 sec.

d. The work done on the body.

e. The kinetic energy as calculated from the final velocity.

3. When an 80-kg. man climbs to a third-story office 15 m. above the street in 3 sec., find the power developed.

4. When a first baseman catches a 150-g. baseball traveling 30 m./sec., his glove moves back 30 cm. Find the average force exerted by the ball through the 30-cm. distance.

5. If at 96 km./hr. the retarding forces of friction on a car are 1,200 newtons, find the power of the engine.

6. One of the alpha particles emitted by radium has a mass of 6.7×10^{-24} g. and a speed of 3×10^9 cm./sec. Find its kinetic energy.

7. Find the kinetic energy of a body weighing 4 kg. after it has fallen freely from rest for 3 sec.

8. A mass of 200 kg. drops 60 m. What is its kinetic energy in ergs and in joules?

9. A small stream has a flow of 100 cu. ft. of water per minute, which goes over a waterfall 40 ft. high. Assume that a power plant is to be installed at the base of the falls. What power in kilowatts can be developed if the installation is 50 per cent efficient? (Mass of 1 cu. ft. of water = 62.5 lb.)

10. The electric locomotives of the Pennsylvania Railroad develop 9,500 hp. How much force can one of these exert in pulling a train 100 mi./hr.?

11. Find the horsepower developed by a boy weighing 110 lb. who, in 4 sec., runs up 15 stair steps, each rising 8 in.

12. The tension in a belt is 90 newtons on one side of a pulley and 20 newtons on the other. If the belt runs 4 m./sec., find the power delivered.

13. An elevator having a mass of 2,500 lb. is lifted 400 ft. in 65 sec. What is the power required? Express this in foot-pounds per second, horsepower, and kilowatts.

14. A man pulls a sled along a level sidewalk, supplying energy at the rate of 100 watts and walking at the rate of 3 m./sec. What horizontal force does he apply to the sled?

15. If a compressed spring can impart a velocity of 100 m./sec. to a 5-kg. mass, find the work that must have been required to compress the spring.

16. A rifle has a mass of 3 kg. A bullet has a mass of 15 g. The muzzle velocity of the bullet is 200 m./sec. Calculate (*a*) the momentum of the bullet, (*b*) the energy of the bullet, (*c*) the momentum of the rifle, (*d*) the kinetic energy of the rifle. Assume in the foregoing that the rifle is free to recoil. How will all these quantities be affected if the rifle is securely fastened to a very massive concrete block?

17. Assume the bullet of Prob. 16 to strike and bury itself in a wooden block having a mass of 2 kg. suspended by a long wire. What will be the velocity of the combined block and bullet immediately after the impact? What will be their kinetic energy? How much energy has been lost in frictional heating? How high will the block rise?

18. A bullet weighing 30 g. imbeds itself in the wooden block of a ballistic pendulum weighing 9.97 kg. and raises it 10 cm. vertically. Find (*a*) the potential energy of the block and bullet at the height of the swing; (*b*) the kinetic energy, velocity, and momentum of the block and bullet just after collision; and (*c*) the kinetic energy, velocity, and momentum of the bullet just before collision. What happens to the energy of the bullet?

19. The work done on an elastic body in stretching it is proportional to the square of the amount of stretch. Show that the force resisting the stretch is proportional to the stretch. (Hooke's law.)

20. The potential energy of position of any mass attracted to the earth is inversely proportional to its distance from the center of the earth. Show that this statement agrees with the inverse-square law of gravitational attraction.

21. A smoke particle in Brownian movement is originally at rest. How many oxygen molecules with velocity of 400 m./sec. would have to collide elastically with it simultaneously in order to give it a velocity of 1 mm./sec.? Assume that the smoke particle is 1.6×10^{12} times as heavy as an oxygen molecule.

22. A cannon of mass 1,000 kg. perfectly free to move on a horizontal plane fires a shell of 10-kg. mass at a 45° angle. If the shell rises to a maximum height of 3,200 m., find the recoil velocity of the cannon.

CHAPTER VI

FRICTION

1. The forces of friction were mentioned briefly in Par. 17, Chap. IV, and also in Par. 25 of the last chapter but have been ignored in the development of the general laws of mechanics. By this omission, it was possible to treat that subject in a completely logical and clean-cut way. Pedagogically, this is pleasant and avoids a great deal of confusion. But if there were no such thing as friction, the world would be too slippery for comfortable living as anyone who remembers his first efforts on skates or skis will acknowledge. There is friction; it is one of the most important forces in the practical world; and it is this world, the world of experience, that we are studying. The very essence of physics is the application of logical analysis and reasoning to the world we observe around us. We do not set up hypotheses about atoms and molecules or formulate the definitions and laws of mechanics merely to exercise our imaginations or develop our powers of abstract logic. This may be valuable but it is not physics. In physics, we must continually keep our feet on the ground of observation and experiment. So let us turn to the earthy and unpleasant subject of friction.

Sliding Friction

2. This is a perfectly familiar phenomenon. Everyone knows that it takes a force to push a book across a horizontal table. Probably almost everyone has noticed that it takes somewhat more force to start the book than to keep it moving after it has been set in motion. It is also common knowledge that the amount of force needed depends on the nature of the surfaces in contact. It takes more force to push a book across a desk blotting pad than across the highly polished surface of the desk itself. Our problem is to make these observations more precise and to interpret them in terms of other knowledge.

3. Evidently we are dealing with the force that resists the motion of one flat surface over another, a force that must be

parallel to the surface of contact. We can imagine two factors that may contribute to this force. First, there must be surface irregularities which press into each other much as the teeth of two cogwheels mesh. The motion of one surface over the other requires the pushing aside of these irregularities or the lifting of one over the other. This may be illustrated by the sliding of two pieces of sandpaper over each other indicated schematically in Fig. 48. The forces between these irregularities of surface are in the last analysis the forces between the molecules forming

FIG. 48.—Greatly magnified schematic diagram of surfaces in contact.

them, the same type of forces responsible for elasticity and surface tension. These forces may still be important even if the surfaces in contact are so highly polished that irregularities are negligible. Forces of attraction between the molecules of one surface and those of the other will still be present and will resist the motion of one surface over the other. These intermolecular adhesive forces are the second important factor contributing to friction. (One might really say that they were the only factor but that they enter both indirectly and directly.) No matter how highly surfaces are polished, it is impossible to eliminate friction between them. Unfortunately from our present knowledge of the structure of surfaces and intermolecular forces, it is quite impossible to make calculations of frictional forces. We have to rely entirely on direct experimental methods.

Law of Sliding Friction

4. Suppose we study the sliding of a rectangular block on the uniform surface of a table. Let the surface be horizontal and the block be moved by a horizontal force supplied by a pan of weights, hanging by a cord over the edge of the table as shown in Fig. 49. We can study the *static friction* by adding weights until the block just begins to slide. To study the *dynamic friction*, we must observe the acceleration of the block for a given force acting. We have the general result that the total force $T - F = ma$ where T is the force applied to the block by the string from the pan of weights, F is the force of friction, m is the mass of the block, and a is the acceleration. F, the force of friction, is a resistance to motion and is therefore opposed to the direction of

the motion. The dynamic frictional resistance is always less than the static so that if the pull of the string is great enough to start the block there will be acceleration. This acceleration can be measured and the frictional force deduced.

5. By observing the acceleration of the block for various weights on the pan, we can find the frictional force for different velocities, and we discover that the sliding friction is independent of the velocity over a considerable range. We shall see later that for many other kinds of frictional resistance this is not true.

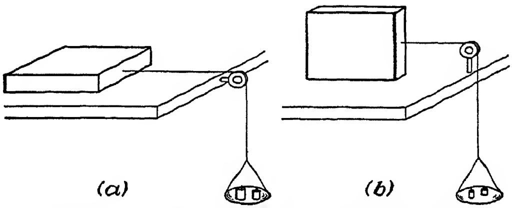


FIG. 49.—A block sliding on a horizontal table.

6. A more surprising result is that the frictional resistance is the same whether the block is on its bottom [Fig. 49(*a*)] or on its side [Fig. 49(*b*)] if the polish of the surfaces is the same. If a weight is put on top of the block, the force of friction is increased by an amount proportional to the weight and independent of whether the block is on its side or bottom. These two results are expressed quantitatively in the following statement known as the Law of Friction.

The force of friction opposing the sliding of one plane surface over another is proportional to the total force pressing the surfaces together and depends on the nature of the surfaces but is independent of the area of contact and the velocity.

Or in mathematical language,

$$F = \mu P \tag{1}$$

where μ is a constant of proportion known as the coefficient of friction and is equal to the ratio of the frictional force along the surface of contact to the force perpendicular to the surface of contact.

7. This law holds for both static and dynamic friction, but the coefficient of static friction is greater than the coefficient of

dynamic friction. Some coefficients of friction are given in Table 3 below. They are so sensitive to small differences in the surfaces that their numerical values are only approximate.

TABLE 3.—COEFFICIENT OF DYNAMIC FRICTION

Glass on glass	0.40	Brass on brass	0.20
Mild steel on mild steel	0.56	Steel on ice	0.02
Hard steel on hard steel	0.42	Rubber on dry concrete	0.5–0.8
Nickel on mild steel	0.65	Leather on metals	0.5–0.6
Carbon on glass	0.18	Leather on metals oiled	0.15
Wood on wood	0.25–0.50	Lubricated metal surfaces	0.05–0.2

The Inclined Plane

8. If the table of Fig. 49 is tilted at an angle θ to the horizontal and there is no string attached to it, the block will tend to slip under the action of gravity. This case is a problem in the resolution of forces which is of practical significance. As in Chap. IV, illustrative problem 3, we can resolve the weight into components parallel and perpendicular to the inclined plane. The perpendicular component is evidently the force perpendicular to the surfaces in contact, the P of Eq. (1). The frictional force F is therefore equal to $\mu P = \mu mg \cos \theta$ directed up the plane. If this is less than the component of the weight down the plane $mg \sin \theta$, the block will slip with an acceleration

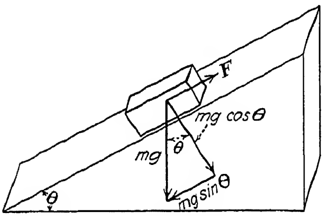


FIG. 50.—A block sliding on an inclined plane.

$$a = \frac{mg \sin \theta - \mu mg \cos \theta}{m} = g(\sin \theta - \mu \cos \theta). \qquad (2)$$

If the acceleration is zero, this equation reduces to $\mu = \tan \theta$. This corresponds to slipping with constant velocity, in which case μ is the coefficient of dynamic friction, or to being just on the point of slipping, in which case μ is the coefficient of static friction. In the latter case, the corresponding angle $\theta = \arctan \mu_s$ is called the angle of repose.

9. This notion of angle of repose is immediately applicable to piles of granular material like dry sand, threshed grain, or small-sized coal. For a given size and kind of granule, there will be a

definite frictional resistance to the motion of the granules over each other. Consequently the pile will spread out with sloping sides which make a definite angle with the horizontal. This angle is a kind of statistical angle of repose.

Rolling Friction

10. The existence of rolling friction depends on the compressibility of all materials. A mathematically perfect sphere rolling on a mathematically perfect plane would make contact with it in a mathematical point and there would be no friction. But a real material ball rolling on a real surface will be somewhat flattened at the point of contact and will make a slight depression in the surface. As the ball rolls, these elastic deformations move and energy is continually used in producing them which is not entirely recovered when they disappear. This effect appears as a frictional moment of force opposing the rolling and is, like sliding friction, proportional to the force across the surface of contact. But the magnitude of rolling friction is usually much less than sliding friction.

Resistance of Fluids to Moving Bodies

11. In these days of high-speed transportation, great importance attaches to the resistance offered by air and water to the motion of cars, airplanes, and boats. For both boats and airplanes, this resistance of fluids is the only important factor controlling their motion, and even for trains and automobiles air resistance becomes very serious at high speeds. The same general laws govern the resistance offered by any fluid, whether it is a gas or a liquid, to the motion of any object through it; and this resistance is essentially frictional in type. But there are two different ways in which this friction manifests itself. In the first, more important at low speeds, there is a slipping of one layer of the fluid over another along surfaces more or less parallel to the direction of motion. In the second type of resistance, higher speeds produce so much disturbance in the fluid that eddies and whirlpools are set up which have considerable kinetic energy that gradually dissipates itself by heating the fluid. Either type of resistance can be studied in terms of the flow of fluids through pipes and around obstacles or in terms of the motion of solid objects through fluids. It is the relative

motion that counts. The first type of resistance is called viscous resistance and the second type turbulent resistance.

Viscosity

12. Everyone is familiar with the differences in the flow of various fluids. Water pours from a pitcher more quickly than honey or maple sirup. "Lighter" oil is used in cold weather than in hot because an oil's resistance to flow increases as the temperature decreases. We call this resistance to flow the viscosity of a fluid and describe a sluggishly flowing liquid like molasses as more viscous than water. This property has nothing to do with the "heaviness" of the fluid in the ordinary sense, *i.e.*, with its density. Water is denser than oil, and mercury, the densest of familiar liquids, is not very viscous. Evidently this is a property akin to surface tension and elasticity, a property depending on the intermolecular forces. These forces resist the motion of one part of a fluid relative to its adjacent parts, and this resistance eventually increases the random motion of the molecules themselves, *i.e.*, develops heat.

13. When a body is moving through a viscous fluid, the molecules on the surface of the body tend to drag along the adjacent

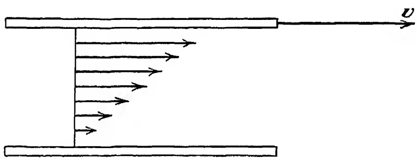


FIG. 51.—Viscous drag between a moving and stationary plate.

molecules of the fluid because of intermolecular forces. These forces in the liquid itself transmit the drag to the next layer of molecules and so on. A certain amount of slipping occurs, and the more viscous the fluid the less is the slipping. A very simple example is shown in Fig. 51, where the upper plate is moving and the lower one is stationary. The arrows in between indicate the direction and magnitude of the motion of successive layers of the fluid. Evidently the viscosity of the fluid transmits a dragging force from the upper plate to the lower. It is found experimentally that the amount of this force depends on the size and separation of the plates, their relative velocity, and the nature

of the fluid. To be precise, if two parallel planes of area A are separated by a thickness d of a fluid, the drag on one plate exerted by the other is

$$F \propto \frac{Av}{d} \tag{3}$$

where v is the relative velocity of the two planes. The constant of proportion necessary to make this an equation is called the coefficient of viscosity and is usually denoted by η (Greek eta). By using η, the proportion in (3) can be written as the equation

$$F = \frac{Av\eta}{d}. \tag{4}$$

The coefficient of viscosity is a constant characteristic of the fluid and is evidently defined exactly as follows:

14. *The coefficient of viscosity of a fluid is the tangential force exerted on a plane surface of unit area by another plane surface of unit area parallel to it and separated from it by unit thickness of the fluid, when the first surface is fixed and the second surface is moving with unit velocity parallel to its plane.*

TABLE 4.—COEFFICIENTS OF VISCOSITY

Substance	Coefficient of viscosity, m.k.s. units*	Temperature, °C.
Air	0.0182×10^{-3}	20
Ethyl alcohol	1.1943×10^{-3}	20
Ethyl ether	0.234×10^{-3}	20
Glycerin	$830. \times 10^{-3}$	20
Hydrogen	0.0097×10^{-3}	20
Mercury	1.56×10^{-3}	20
Oil, castor	$986. \times 10^{-3}$	20
Oil, machine	$100\text{–}600 \times 10^{-3}$	20
Oil, olive	$84. \times 10^{-3}$	20
Water	1.7938×10^{-3}	0
	1.0087×10^{-3}	20
	0.2839×10^{-3}	100

* The m.k.s. unit of viscosity is 1 newton-sec./sq. m. This unit is ten times as large as the c.g.s. unit, the dyne-second per square centimeter.

15. In discussing the highly simplified experiment of Fig. 51, we stated that the amount of drag depended on the relative

velocity of the plates. This is found experimentally to be a general result for the resistance of a viscous medium to a body moving through it as long as the speed is so low that there is no turbulence. Furthermore, the nature of the material in the body makes no difference. Apparently the first layer of the molecules of the fluid cling to the surface of the body without moving over it so that the only slipping is between successive layers of the fluid itself.

16. The definition of coefficient of viscosity and the other remarks we have made apply equally well to liquids and gases although the numerical values for the coefficient are very different, as can be seen from the few typical values given in Table 4. We shall have occasion to apply these ideas later on in Chap. XII when we are considering the charge of an electron.

Lubrication

17. If a film of liquid can be interposed between two solid surfaces, the frictional resistance to the motion of one solid relative to the other is reduced. If the film is thick enough, the resistance will depend only on the viscosity of the liquid. Each solid will carry a thin film of liquid along with it, and all the slipping will be between layers of the liquid itself. We have defined the coefficient of friction and the coefficient of viscosity in such different ways that a direct comparison of the numbers of Table 3 with those of Table 4 is meaningless. But it may be interesting to calculate the effect of a film of oil for a particular case.

18. Suppose we have a cube of brass 10 cm. along each edge resting on a horizontal brass plate. The coefficient of friction as we see from Table 3 is about 0.2, and the density of brass is about 8,000 kg./cu. m., or 8 g./cc. Therefore the force across the surface of contact is $mg = 8 \times 9.8 = 78.4$ newtons, and the frictional force is 0.2 times this, or 15.7 newtons. In order to compare this with the viscous drag of a film of lubricant, let us assume that a film 0.1 mm. $= 10^{-4}$ m. thick of castor oil is between the surfaces and that the cube of brass is moving 5 cm./sec. or 5×10^{-2} m./sec. Substituting these values and the coefficient of viscosity for castor oil from Table 4 in Eq. (4), we get 4.93 or approximately 5 newtons. Evidently we have reduced the friction to about one-third. This checks roughly with the direct

determination of the coefficient for greased surfaces given in Table 3.

19. Obviously the coefficient of viscosity is not the only important factor in choosing a lubricant. The film of lubricant must not dry up or decompose when hot or be forced out by the pressure of the bearing. If the film of liquid is very thin, the simple law of viscous drag is no longer applicable, the friction is less than without any lubricant but behaves more like dry friction than like fluid friction.

Plastics

20. Many materials that are not ordinarily considered fluids can be made to flow under great pressure acting for long times. Pitch and wax are such materials and so is the ice in a glacier. Even metals can be made to flow by sufficiently high pressure. This is essentially the process used in stamping out many shaped pieces of sheet metal. The coefficients of viscosity for such materials are many thousand times those given in Table 4.

Turbulent Resistance

21. When the speed of a body moving through a fluid medium increases beyond a certain point, the resistance begins to increase more rapidly than the velocity. It depends on the square of the velocity for certain ranges of speed but then may increase somewhat less rapidly as very high speeds are approached. The exact relation is impossible to predict theoretically and difficult to determine experimentally. At least one thing is clear. As the speed of a body is gradually increased from the region of purely viscous resistance, there comes a time when eddies and vortices begin to develop in the fluid and this general disturbance of the fluid corresponds to greatly increased resistance. The amount of turbulence and the speed at which it sets in depend very much on the shape of the body. If a body is "streamlined," the effect of turbulence is very much reduced.

22. A detailed discussion of streamlining is obviously beyond the scope of this book, but we shall describe a simple example because of the very general interest in the problem at the present time. Consider the case of a cylindrical body such as is shown in Fig. 52(*a*), and assume that it is moving through air with a

velocity in the range where the resistance is proportional to the square of the velocity. This resistance is the result of pressure on the front of the disk and suction on the back which sets up turbulence. This air resistance, or air drag as it is often called, can be reduced without altering the cross section of the body perpendicular to its direction of motion. The amount of the reduction for different changes in shape is shown in the drawings in Fig. 52. The resistance of the original shape is taken as unity.

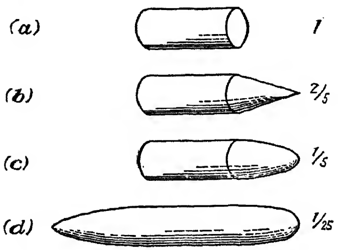


FIG. 52.—Effect of streamlining. The differently shaped bodies are all moving to the right. The numbers give their relative resistances to motion.

Terminal Velocity

23. According to Newton's second law of motion, the constant application of force to a body must continue to increase its velocity, but we know that any moving object whether it be runner, train, or airplane soon reaches a maximum speed even if energy continues to be supplied. The explanation of this paradox is apparent from the foregoing discussion. The moving system continues to accelerate until the increasing resistance of friction is just equal to the maximum accelerating force supplied by the machine. At this point, the resultant force on the system is zero and Newton's first law is satisfied. If frictional resistance did not increase with speed, there would be no limit to the speeds attainable by the prolonged application of even moderate power. The velocity finally reached by any system accelerated by a constant applied force is called the *terminal velocity* of the system. Perhaps the simplest example is that of a body falling through the atmosphere under the constant force of gravitation. Whether the air resistance varies as the first or second or any other power of the velocity of the falling body is immaterial; just so long as it does increase with velocity, the falling body will eventually move with constant speed. What this speed is will depend on the shape and average density of the falling body. The terminal velocity of a man falling from an airplane is about 120 mi./hr. If he opens his parachute, this is reduced to about 12 mi./hr., which gives an impact with the ground equivalent to a jump from a height of about 5 ft.

SUMMARY

No motion occurs without some friction. Frictional forces arise from intermolecular forces. The simplest type of friction is that between two flat surfaces in contact. A distinction must be made between static friction, the force resisting the beginning of motion, and dynamic friction, the resisting force that continually retards motion. In each case, the frictional force is along the surface of contact and proportional to the total force perpendicular to this surface.

Rolling friction results principally from elastic distortion and is much less than sliding friction.

The internal friction that retards the motion of one part of a fluid relative to another is called viscosity. The resistance of air and water to bodies moving through them arises from viscosity. There are two kinds of fluid resistance, turbulent and nonturbulent. Because viscous resistance increases as the velocity increases, a body under the action of a constant applied force reaches a terminal velocity in a fluid.

ILLUSTRATIVE PROBLEMS

1. A man pulls a sled weighing 50 kg. by a rope making an angle of 30° with the horizontal. What force must he exert in order to maintain a constant speed if the coefficient of friction between the sled and the snow is 0.05?

The force of friction that acts in the direction opposite to the motion of the sled is equal to the coefficient of friction times the resultant normal component of the forces holding the sled on the snow. This normal component is the weight of the sled minus the vertical lift of the man's pull. This latter force in Fig. 53 is $F \sin \theta$ where F is the pull of the man on the rope. Thus the force of friction is

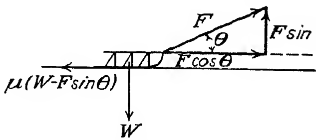


FIG. 53.—Forces acting on a sled.

$$f = \mu(W - F \sin \theta)$$
$$f = 0.05(50\ g - F \sin 30°).$$

This force must be exactly balanced by the component $F \cos 30°$ in the direction of motion of the man's pull. Therefore,

$$f = 0.05(490 - F \sin 30°) = F \cos 30°,$$

and

$$F \cos 30° + 0.05\ F \sin 30° = 0.05 \times 490 \text{ newtons}$$
$$(0.866 + 0.025)F = 24.5 \text{ newtons}$$
$$F = 27.4 \text{ newtons}$$

2. A force of 500 newtons is applied at an angle of 30° to an inclined plane making an angle of 15° with the horizontal to drag a 50-kg. block up the plane. With what acceleration will the block move up the plane if the coefficient of sliding friction between the block and the plane is 0.4?

Three forces act on the block (Fig. 54) the applied force, the weight of the block, and the reaction of the plane. The reaction of the plane consists of two components: one perpendicular to the plane and the force of friction parallel to the plane. The component of the reaction perpendicular to the plane balances the components of the pull and the weight perpendicular to

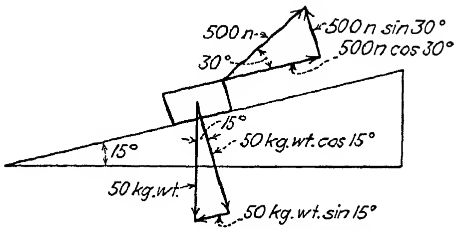


FIG. 54.—Forces acting on a block on an inclined plane.

he plane. The force of friction is equal to the coefficient of friction times his perpendicular reaction and acts down the plane. The perpendicular reaction is thus

$$\begin{aligned} R &= 50 \text{ kg.} \times 9.8 \text{ newtons/kg.} \times \cos 15^\circ - 500 \text{ newtons} \times \sin 30^\circ \\ &= 50 \times 9.8 \times 0.966 \text{ newtons} - 500 \times 0.5 \text{ newtons} = 223 \text{ newtons.} \end{aligned}$$

The force of friction is

$$f = \mu R = 0.4 \times 223 \text{ newtons} = 89 \text{ newtons.}$$

In addition to the force of friction down the plane, the weight of the block has a component down the plane of

$$\begin{aligned} 50 \text{ kg.} \times 9.8 \text{ newtons/kg.} \times \sin 15^\circ &= 50 \times 9.8 \times 0.259 \text{ newtons} \\ &= 127 \text{ newtons.} \end{aligned}$$

Also the applied force has a component of

$$500 \text{ newtons} \cos 30^\circ = 500 \times 0.866 \text{ newtons} = 433 \text{ newtons}$$

up the plane. The resultant force that will accelerate the block up the plane is therefore

$$433 \text{ newtons} - 127 \text{ newtons} - 89 \text{ newtons} = 217 \text{ newtons.}$$

The acceleration of the block produced by this force is

$$a = \frac{F}{m} = \frac{217 \text{ newtons}}{50 \text{ kg.}} = 4.34 \text{ m./sec.}^2$$

PROBLEMS

1. If in Fig. 55 the coefficient of starting friction is 0.4, find the force in the direction F necessary to start the 100-kg. block in motion.

2. If at 100 km./hr. (62.2 mi./hr.) the retarding forces of friction on a car are 100 kg. weight, find the power of the engine.

3. A block A rests on another block B which in turn rests on a table top. The coefficient of starting friction between A and B is 0.2. How much acceleration must a force, applied to B, give B in order to jerk B out from under A?

4. A block having a mass of 10 kg. rests on a horizontal surface where the coefficient of starting friction between them is 0.5 and the coefficient of sliding friction is 0.4. If a constant force just sufficient to start the block is applied to it, find the acceleration.

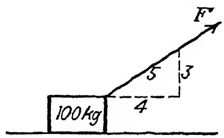


FIG. 55.—A force F acting on a 100-kg. block. Prob. 1.

5. A man on skis weighs 80 kg. If he is sliding down a 30° slope on which the coefficient of friction is 0.10, find (*a*) the force of friction holding him back, (*b*) his acceleration, (*c*) his velocity after sliding 10 sec., and (*d*) the distance he will go on a level with this initial velocity and the same coefficient of friction.

6. To keep himself from starting down a 45° slide with a starting coefficient of 0.15, an 80-kg. skier holds onto a rope. Compare the forces in the rope when it is horizontal and when it is parallel to the slide.

7. A force of 5.9 newtons acts on a block weighing 1 kg. resting on a horizontal table. The coefficient of sliding friction between the block and the table is 0.5. What is the acceleration of the block?

8. Find the force necessary to give a block weighing 32 kg. an acceleration of 1 m./sec.^{2} on a horizontal plane where the coefficient of sliding friction is 0.1.

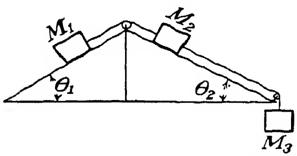


FIG. 56.—Masses on a double inclined plane. Prob. 10.

9. Find the force parallel to the plane in Prob. 8 if the plane is inclined at 30°, the acceleration being (*a*) up the plane and (*b*) down the plane.

10. Three masses M_1, M_2, and M_3 are connected by a weightless string. M_1 and M_2 rest on inclined planes making angles θ_1 and θ_2 with the horizontal while M_3 swings free, Fig. 56. (*a*) Neglecting friction, what is the acceleration of the system? (*b*) What is the tension in the string between M_3 and M_2? (*c*) What must be the coefficient of friction in order that the acceleration be zero?

11. A boy runs up a hill dragging a 10-kg. sled by a rope that makes an angle of 30° with the ground. The angle of the hill is 30°, the coefficient of friction is 0.2, and the acceleration of the sled 0.5 m./sec.^{2} What is the tension in the rope?

12. The viscosity of liquids may be measured by rotating one cylinder inside a fixed coaxial cylinder. The length of the cylindrical film of oil

between the cylinders is 20 cm., its thickness is 0.1 mm., and the radius of the cylinders is 10 cm. The inner cylinder makes 600 r.p.m. and the torque opposing rotation is 79 newton-m. What is the viscosity of the oil? Consider each cylinder a succession of small plane strips, and neglect the effect of the difference in radius of the two cylinders on the torque.

13. A 50-kg. parachute jumper waits 11 sec. before pulling his rip cord. If the rate of fall with an open chute is 9.80 m./sec. and it takes 1 sec. to slow down to this velocity, what is the average force exerted by the chute on the jumper? What is the force after the deceleration is over? Neglect air friction before the chute opens.

CHAPTER VII

TEMPERATURE AND HEAT

1. In the first two chapters, we discussed the nature of matter in a qualitative way talking about the different kinds of atoms, the forces between them, their motions, and so on. It became clear that we had to know more about force and motion before we could deal more precisely with such problems. In the last four chapters, we have discussed the principles of mechanics not only in order to be able to apply them to molecular phenomena but because they are fundamental to the whole of physics. We are by no means confining ourselves in this book to molecular physics. We are merely using the molecular interpretation of phenomena as the coordinating theory which gives coherence to the whole of modern physics. Bearing this in mind, we shall now turn to the subject of heat. We have already suggested that temperature is connected in some way with rapid chaotic motion of molecules. The chapters on mechanics show how motion and energy are to be handled, but before we can apply these principles to a molecular interpretation of temperature we must review the general subject of temperature and heat.

Temperature

2. In beginning the study of heat, it is not necessary to introduce entirely unfamiliar ideas but to give precision to old ones. Everyone has some notion of temperature, of the difference between a hot day and a cold day, and could even arrange a series of hot bodies in order of ascending temperature. Most people even know in a general way how the temperatures of these bodies could be measured so that they might be compared quantitatively. It is with a consideration of the measurement of temperature that this chapter on heat will begin.

Thermometers and Temperature Scales

3. For ordinary temperature ranges, say from -40 to $+212$°F. (*i.e.*, from -40 to $+100$°C.), thermometers are used which depend on the fact that substances change their volume as the temperature changes, *i.e.*, they expand or contract with rising and falling temperatures. Most substances increase in volume as the temperature rises. A notable exception is water which contracts slightly as the temperature is raised from the melting point to 4°C. and then expands for a further rise in temperature. For this reason, water is not a good liquid to use in a thermometer even for measuring temperatures above the freezing point. The liquids most commonly used for filling thermometers are alcohol and mercury.

4. The way in which a mercury-in-glass thermometer is calibrated is typical of the procedure followed with all thermometers and shows clearly the nature of the scale of temperature. The thermometer is allowed to stand in melting ice at normal atmospheric pressure until it has taken up the temperature of the ice bath; a mark is then made on the tube at the position of the top of the mercury column. The thermometer is next brought to the temperature of boiling water again at normal atmospheric pressure and another mark made at the new position of the end of the column of mercury. The first mark is labeled 0 and the second one 100 for a centigrade scale, the space on the tube between them being divided into 100 equal divisions and labeled from 0 to 100. The quantitative measure of any intermediate temperature is then the reading of the division where the end of the mercury column is found. A temperature of 50° is thus the temperature at which the mercury has a volume half way between its two volumes at the temperature of melting ice and boiling water, respectively (the effect of the slight expansion of the glass of the thermometer being neglected). The temperature scale obviously must depend largely upon the nature of the thermometric substance used. Some liquid other than mercury, say toluene, could be used, and the process of calibration would be the same. A temperature that would be 50° on the mercury thermometer would presumably be somewhat different on the toluene thermometer, for there is no reason to expect that the two liquids, mercury and toluene, should both expand with

rising temperature in the same way. The toluene might reach a volume half way between its volume at 0°C. and its volume at 100°C. at a temperature different from that at which the mercury reached a corresponding volume.

5. Any other property, such as electrical resistance, can be used for measuring temperature in the same way. In general, the numerical measure of the temperature on a centigrade scale will be given by the expression

$$t = \frac{p_t - p_0}{p_{100} - p_0} \times 100 \tag{1}$$

where p_0, p_{100}, p_t are the values of the property of the thermometric substance at the temperatures of melting ice, boiling water, and at the temperature being measured. It should, thus, be clear that there is no reason whatsoever why temperatures measured on the scales of different types of thermometers should agree even approximately except at the arbitrarily fixed temperatures of 0 and 100°. Every such temperature scale depends very largely on the nature of the thermometric substance employed. For example, a water thermometer calibrated by the method just described would give a small minus reading at a temperature that a mercury thermometer would record as +4°C. The size of a degree and the temperature to be called zero is entirely arbitrary on all scales ordinarily used as may be seen from the wide differences between the Fahrenheit, Réaumur, and centigrade systems. The centigrade degree is always used in scientific work, though we shall see later that it is sometimes more convenient to count the number of degrees from the "absolute zero" (−273°C.) than from the melting point of ice (0°C.).

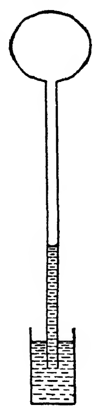


FIG. 57. Galileo's gas thermometer.

6. The thermometers that use a gas as the thermometric substance merit special consideration since it has been found possible by studying their behavior to define a scale of temperature more fundamental than that of any single thermometer.

7. As an example of a crude gas thermometer, we may cite that used by Galileo. It consisted of a bulb and an open-ended glass tube. The mouth of the tube was submerged in water as shown

in Fig. 57, and some of the air was forced out so that the water rose in the tube as shown. The height at which the water stood thereafter depended on the temperature and therefore served to measure it. But it also depended on the barometric pressure and therefore gave results of little accuracy.

8. A gas can be used as a thermometric substance in two different ways, either the pressure to which it is subjected can be maintained constant and the increase of its volume with rising temperature measured, or else its volume can be kept constant and the rise in pressure measured.* We shall confine our discussion to the first case as it is analogous to the mercury or alcohol thermometer.

9. A gas thermometer, then, will be calibrated according to the procedure outlined in Par. 4 where the property to be observed is the volume, the pressure being maintained constant. We have then the temperature t given by

$$t = \frac{V_t - V_0}{V_{100} - V_0} \times 100 \tag{2}$$

where V_t is the volume at the temperature being measured, V_0 that at the melting point of ice, and V_{100} that at the boiling point of water. The quantity

$$\frac{V_{100} - V_0}{100 V_0}$$

is the average change in volume per unit volume per degree and

* In 1887, the International Committee on Weights and Measures agreed on the following method of measuring temperature:

"Temperature shall be measured by the gaseous pressure of a constant volume of chemically pure hydrogen, which at the melting-point of ice exerts a pressure equal to that of a mercury column 1,000 mm. in length. The fixed points of the thermometer shall coincide with those of the Celsius (centigrade) scale."

This empirical definition of a temperature scale corresponds to the definition of the kilogram and meter in terms of specific pieces of metal at the International Bureau of Weights and Measures. In practice, either the constant-volume or the constant-pressure hydrogen thermometer departs slightly from the thermodynamic or perfect gas scale. Helium thermometers, either constant pressure or constant volume, are better but helium was unknown in 1887. The corrections to be applied to either helium or hydrogen thermometers to reduce their readings to equivalent temperatures on the perfect gas or thermodynamic scale can be found in handbooks such as the *Handbook of Chemistry and Physics* published by The Chemical Rubber Company.

should be independent of the amount of gas used. It is called the temperature coefficient of expansion and may be written as α. Then Eq. (2) becomes

$$t = (V_t - V_0) \times \frac{100}{V_{100} - V_0} \times \frac{V_0}{V_0} = \frac{V_t - V_0}{V_0} \times \frac{1}{\alpha} \qquad (3)$$

We are now in a position to appreciate the advantages of a gas thermometer, for experiments show that

1. The coefficients of thermal expansion α for different gases, though differing somewhat at pressures as high as one atmosphere, all approach almost exactly the same limit as the pressure and the density of the gas are decreased.

2. Not only do the coefficients representing the *average* expansion between 0 and 100° approach the same limits, but thermometers filled with different gases approach the same reading as the pressure decreases for any intermediate temperature.

10. These results are of such importance that it is worth while to quote some experimental data. Some observations of Henning and Heuse* give the following results:

COEFFICIENTS OF VOLUME EXPANSION AT CONSTANT PRESSURE

Gas	Pressure, m. Hg	α, $\frac{\text{cu. m.}}{\text{cu. m. deg.}}$
Helium	1.12	36.581×10^{-4}
	1.10	36.582×10^{-4}
	0.76	36.591×10^{-4}
	0.52	36.603×10^{-4}
	0.50	$36.589? \times 10^{-4}$
	0	36.600×10^{-4}
Hydrogen	1.10	36.590×10^{-4}
	0.51	36.602×10^{-4}
	0	36.607×10^{-4}
Nitrogen	1.11	36.743×10^{-4}
	0.51	36.679×10^{-4}
	0.22	36.630×10^{-4}
	0	36.606×10^{-4}

The values for zero pressure are, of course, found by extrapolation, that is, by plotting the observed points on a graph and

* *Zeitschrift für Physik*, 5, 285–314 (1921).

then drawing a smooth curve through them and on down to zero pressure. It is clear that the limiting values are very nearly the same for all three gases. The same value is also found for other gases even such as oxygen or carbon dioxide which have considerably different values of α at atmospheric pressure.

11. For our purposes, it will be sufficiently accurate to take as the limiting value $\alpha = \frac{1}{273} = 36.630 \times 10^{-4}$, and we therefore have for all gases, by rearrangement of Eq. (3),

$$\alpha t = \frac{V_t}{V_0} - 1$$

or

$$V_t = V_0\left(1 + \frac{t}{273}\right) \tag{4}$$

where V_0 is the volume at temperature 0°C. and V is the volume at temperature t measured, strictly speaking, by a gas thermometer but equally well in practice by a mercury one. This is known as the Law of Charles and Gay-Lussac. It may be restated as follows:

All gases have very nearly the same coefficient of thermal expansion. In terms of the centigrade degree and volume at zero degrees centigrade, this coefficient is approximately $\frac{1}{273}$. Therefore the volume V_t of a gas at any temperature t is given by

$$V_t = V_0\left(1 + \frac{t}{273}\right)$$

where V_0 is the volume at zero degrees centigrade.

12. Besides the scale of temperature given by the ideal gas thermometer, there is another closely related and theoretically more important scale known as the absolute thermodynamic scale. It is based on abstract ideas which we shall not consider but it brings us to the idea of an absolute zero of temperature which we can also get from a little further discussion of Charles' law and the gas thermometer.

13. Suppose we consider the effect of large changes in temperature on the volume of a certain amount of gas kept under constant pressure. To be specific, let us think of the gas as confined in a long cylinder of unit cross section in which there is a gastight frictionless piston. Suppose the force on the piston

is kept constant so that the pressure on the gas is constant but that the temperature of the cylinder can be varied by immersing it in various liquids such as liquid air, cold or boiling water, etc. Then the volume of the gas will vary with temperature according to Charles' law as we have just explained. If we neglect the small expansions and contractions of the glass container, the volume of the gas will be measured by the position of the piston. As the temperature of the gas increases, the piston will rise; and as the temperature falls, the piston falls with it. This is shown diagrammatically in Fig. 58. Suppose we call the bottom of the tube A, the position of the piston at 0°C. B (this will be the position the piston will take if the whole tube is immersed in a mixture of ice and water), and the position of the piston at 100°C. C (this is the position the piston will take if the tube is immersed in steam over water boiling at atmospheric pressure). According to our previous discussion, we can divide the distance BC into a hundred parts and we then define the temperature of any bath that makes the piston stand in between B and C in terms of these divisions. We can also extend this temperature scale upward beyond C. Suppose for example that the tube is immersed in boiling glycerin. We find that the piston stands at D and that the distance BD is equal to 2.91 times BC. Therefore we say that glycerin boils at a temperature of 291°C., and we divide up the length CD into 191 "degrees" of the same size as those between B and C. We can continue these graduations upward indefinitely and measure higher temperatures in terms of them.

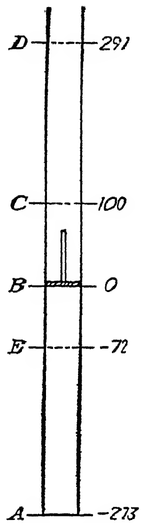


FIG. 58.—Determination of a temperature scale by the expansion of a gas.

14. Just to make sure that this is in agreement with our statement of Charles' law in Eq. (4), let us consider the situation from that point of view. The V_t in that equation is the volume at some particular temperature t and therefore corresponds to the distance from A to some particular position of the piston in Fig. 58. The V_0 is the volume at 0°C. and it therefore corresponds to AB. If t is 100°C., V_t is AC but from (4) $V_t = V_0\left(\frac{373}{273}\right)$. Therefore $AC = AB\left(\frac{373}{273}\right)$. Again if t is

291 centigrade, $V_t = V_0\left(\frac{564}{273}\right)$ and $AD = AB\left(\frac{564}{273}\right)$. Similarly any other temperature as read on this gas thermometer will be consistent with Charles' law if the graduations of the thermometer are each equal to $\frac{1}{273}$ of the distance AB. There is no particular difficulty about this. What we are really doing is to carry out in detail the establishment of a temperature scale discussed in a more general way in the earlier paragraphs of the chapter. But so far we have established the scale only for temperatures above zero. We want next to consider what happens at low temperatures.

15. The most obvious procedure for extending our temperature scale toward temperatures below zero is merely to start marking off divisions below B equal to those above it. We test this procedure just as before by taking a particular example. Suppose that the tube is buried in dry ice and alcohol. The piston will come to a position E, and we see that BE is $\frac{72}{100}$ of BC. This again is perfectly consistent with Eq. (4) if we take t as -72. We need not take other examples. Let us go on to consider what happens to the volume of the gas as we go to lower and lower temperatures. For each degree of temperature, the volume diminishes by $\frac{1}{273}$ of its amount at zero. This process cannot continue indefinitely. In fact, it can only continue to a temperature of -273°C. because at that point the volume of the gas will have shrunk to nothing. Beyond that point, Eq. (4) would give negative volumes which obviously are meaningless.

16. Now as a matter of fact, no real gas obeys Charles' law down to the point of vanishing volume though helium comes pretty close to it. All real gases liquefy at sufficiently low temperatures and solidify at still lower temperatures. It is nevertheless useful to use the point A of Fig. 58 as a reference point from which to measure temperature. It is called the absolute zero of temperature, and temperatures measured from this point are called absolute temperatures. We shall see later when we discuss the kinetic theory of gases in detail that the motion of the molecules diminishes with decreasing temperature and that the absolute zero is the degree of cold where all molecular motion has practically stopped. Since we cannot have less than no motion, we cannot have a temperature lower than the absolute zero. We shall also see in that later discussion why the zero of the gas thermometer comes at this same point. Perhaps it is

already qualitatively evident that the volume of a gas will approach zero as the motions of the molecules which tend to keep them apart approach zero. As we mentioned in Par. 12, another way of establishing this absolute scale of temperature and perhaps the way that is most satisfactory logically is by thermodynamic considerations but these are beyond the scope of this book.

17. One advantage of using absolute temperatures is immediately apparent if we rewrite Charles' law in terms of degrees absolute (denoted by °K. after Lord Kelvin) instead of degrees centigrade. The temperature in degrees absolute equals the temperature in degrees centigrade plus 273 or, in the usual symbols, $T = t + 273$ or $t = T - 273$. Substituting this in equation (4), we get

$$V_T = V_0\left[1 + \frac{T - 273}{273}\right] = V_0\frac{T}{273} \tag{5}$$

which is a considerably simpler expression for the volume than Eq. (4). Using (5) and putting in V_1 and V_2 for the volumes at two different absolute temperatures T_1 and T_2, we get

$$V_1 = V_0\frac{T_1}{273}, \qquad V_2 = V_0\frac{T_2}{273},$$

or dividing, we get

$$\frac{V_1}{V_2} = \frac{T_1}{T_2} \tag{6}$$

which gives us the volume at any temperature in terms of the volume at any other temperature. It is to be remembered that throughout this discussion the pressure is supposed to be kept constant. The effect of pressure variations will be considered in Pars. 22 to 25 below.

18. By means of liquid air, temperatures as low as $-180°$C. may be obtained; and by the evaporation of liquid hydrogen, $-258°$C. has been reached, only 15° above the absolute zero. By the evaporation of liquid helium at low pressure, a temperature of only 0.8° above the absolute zero has been reached. Within the last few years still lower temperatures, certainly less than 0.1° absolute, have been reached by using the cooling effect of demagnetization in addition to preliminary cooling with liquid helium. At these low temperatures, rubber and steel

become as brittle as glass, lead becomes stiff and elastic, and the electrical resistances of metals are greatly reduced.

Thermal Expansion in General

19. Now that a temperature scale has been established in terms of the expansion of an ideal gas, we can discuss the general question of thermal expansion of solids and liquids. It is found experimentally that the thermal expansion of substances can be precisely and concisely described by the equation

$$V_t = V_0(1 + \alpha t + \beta t^2 + \gamma t^3) \tag{7}$$

where V_0 is the volume at 0°C., V_t the volume at t°C., and α, β, and γ are coefficients characteristic of the substance. For most substances, β and γ are so much smaller than α that the βt^2 and γt^3 terms of Eq. (7) may be neglected for any but the most refined work. In particular, this is true for mercury where

$$\alpha = 0.18 \times 10^{-3}, \qquad \beta = 0.00078 \times 10^{-6}, \qquad \gamma = 0$$

which shows why readings on a mercury thermometer agree very closely with the gas scale. The coefficient α for alcohol is 1.04×10^{-3} and for glass it is 0.013×10^{-3}, showing why the expansion of the glass container of a thermometer can ordinarily be neglected. In contrast to mercury, water has the values $\alpha = -0.0643 \times 10^{-3}$, $\beta = 8.505 \times 10^{-6}$, $\gamma = 6.790 \times 10^{-8}$ where the negative sign of α gives the contraction between 0° and 4° mentioned before.

TABLE 5.—COEFFICIENTS OF CUBICAL THERMAL EXPANSION
$V_t = V_0(1 + \alpha t + \beta t^2 + \gamma t^3)$.

Substance	Temperature range, °C.	α	β	γ
Ethyl alcohol	0–80	1.0414×10^{-3}	0.7836×10^{-6}	1.7168×10^{-8}
Mercury	24–299	0.18182×10^{-3}	0.00078×10^{-6}	
Pyrex glass	21–471	0.0108×10^{-3}		
Water	0–33	-0.0643×10^{-3}	8.505×10^{-6}	6.790×10^{-8}

20. Sometimes when a rigid body is concerned, the expansion along one dimension is of more interest than the volume expansion. For example, the length L of a metal bar such as a rail or girder is given by

$$L_t = L_0(1 + at) \tag{8}$$

where L_0 is the length at 0°C. and a is the so-called coefficient of linear expansion. It is easy to show that for any given substance a is approximately equal to $\frac{1}{3}\alpha$ where α is the coefficient of volume expansion. The linear expansion of a metal rod can be used to measure temperature and is so used in many thermostats and recording thermometers.

TABLE 6.—COEFFICIENTS OF LINEAR THERMAL EXPANSION
Change in length per unit length per degree Centigrade,
i.e., $L_t = L_0(1 + at)$

	a in $\frac{\text{m.}}{\text{m.°C.}}$
Aluminum	2.394×10^{-5}
Brass	1.818×10^{-5}
Copper	1.771×10^{-5}
Glass, flint	0.78×10^{-5}
Hard	0.97×10^{-5}
Pyrex	0.3×10^{-5}
Soft	0.85×10^{-5}
Gutta-percha	19.8×10^{-5}
Platinum	0.89×10^{-5}
Platinum-iridium	0.87×10^{-5}
Quartz	0.058×10^{-5}
Steel	1.332×10^{-5}
Tungsten	0.444×10^{-5}
Wood along the grain, mahogany	0.3×10^{-5}
Pine	0.5×10^{-5}
Wood across the grain, mahogany	4.0×10^{-5}
Pine	3.4×10^{-5}

21. In the second chapter it was shown that the molecules of a gas, a liquid, or even a solid are constantly in motion. This fact can be used to give an immediate qualitative explanation of the experimental results we have been discussing if we assume that the motion of the molecules increases with the temperature. The pressure of a gas on a surface must depend on how many molecules of the gas hit the surface every second and on how hard they hit. If the volume of the gas is kept constant, then the number of molecules in a given volume must remain constant and increased motion of the molecules manifests itself in an increase of pressure. But this can be compensated if the volume is allowed to increase so that the number of molecules hitting a given surface decreases. Similarly in a solid or liquid, increased motion of the molecules may be expected to increase the average distance between the molecules, causing expansion, but here the

phenomena would be more complex because of the effects of intermolecular forces. As we have seen, these forces can be neglected in gases at low pressures and therefore we shall find it possible to give an exact quantitative interpretation of the thermal expansion of gases in the next chapter.

Boyle's Law

22. In discussing the variation of the volume of a gas with the temperature, we have been careful to specify that the pressure of the gas was maintained constant and preferably small. But if we examine our general assumptions about the nature of matter in order to get a qualitative explanation of the volume changes of a gas with temperature, we find that pressure changes may also be expected to occur. The pressure that a gas exerts on the walls of the vessel which contains it is, on our theory, merely the result of the continual bombardment of the millions of molecules of which the gas is made up. The amount of this pressure will certainly depend on both the number of molecules hitting the walls and on the vigor of their motion. If the latter increases with temperature, then the pressure must increase with temperature unless the former is decreased. Now pressure is, by definition, force per unit area. We measure pressures in pounds weight per square inch or in newtons per square meter. Therefore, if the volume of a gas is increased, the area of the walls of the vessel containing it must increase and the average number of molecules hitting unit area of the walls must decrease, since the total number of molecules in the vessel remains the same. It is this effect then that made it possible to carry out the experiments we have been discussing where the pressure was kept constant by allowing the increase in volume to compensate the increase in molecular motion caused by rising temperature.

23. We now wish to investigate how the pressure of a gas varies. There are two ways in which we might proceed. One would be to study the variation of the pressure with the temperature, keeping the volume constant. This is perfectly practicable experimentally. It could be done with the apparatus of Fig. 58. if we kept the piston in a fixed position and attached some device for measuring the pressure. We would find results exactly analogous to those of Par 11. We would find that the pressure

at constant volume increased according to the same law as the volume at constant pressure, and we could then combine the two results in one general law. It is more illuminating and more interesting historically to proceed by a different path to the same result.

24. We are dealing here with three variables; volume, pressure, and temperature. Apparently there are relations between them. It is characteristic of the method of experimental science to bring out these relations by studying them piecemeal. It is the constant effort of the experimental scientist to control his experiments so that there are only two things changing at once. He can then correlate the change of one with the change in the other. He is happiest if this correlation takes a simple mathematical form expressing one variable as a function of the other, but this is not essential to the method. Thus in the study of the behavior of gases the three variables we have just mentioned should be considered in pairs. We have discussed how the relation between temperature and volume is determined in an experiment where the pressure is not varied. We can perform two other sets of experiments. In one, the volume will be kept constant and the variation of pressure with temperature observed; in the other, the temperature will be kept constant and the variation of volume with pressure observed. It will not be necessary to do both since either one combined with the volume-temperature result must give the other. That is, if we have three mutually dependent variables A, B, and C and find the relation between A and B and between A and C, we can predict the relation between B and C. In the next paragraph, we shall report the experiments performed by Robert Boyle in 1660 on the relation between the pressure and volume of a gas when the temperature is kept constant.

25. The apparatus that Boyle used is shown schematically in Fig. 59. It is essentially the same as the apparatus used today in elementary laboratory courses to study the same effect. It is simply a J-shaped glass tube with one end A closed and the other end open. Some mercury is poured in, trapping some air in the closed part of the tube AA'. By tilting the tube back and forth, air bubbles can be made to go in and out of AA' until the air there is at exactly atmospheric pressure and the mercury in the two tubes is consequently standing at the same height. Mercury

is then added carefully with the tube standing vertical. The air in AA' is compressed to a smaller volume which is measured by the distance AA'' from the closed end to the top of the mercury. The pressure is then equal to the difference in height of the mercury in the two tubes plus the weight of the atmosphere resting on the mercury in the open side of the tube. This could be read on a barometer, and in Boyle's experiments was equivalent to $29\frac{1}{8}$ in. of mercury. Therefore the total pressure on the air in AA'' is proportional to the distance BB' measured in inches plus $29\frac{1}{8}$. The results that Boyle obtained by putting more and more mercury in the tube are given in the table on page 149. These figures are taken directly from his table except that the first column gives the volume explicitly which he did not bother to do. Each number in it is the difference of two numbers in his table. The second column is the observed pressure as explained above. The last column is of particular interest. It is the value that Boyle calculated the pressure should have if it were inversely proportional to the volume as he saw was indicated by his experiments. In other words, the agreement between the last two columns showed how closely the experiments proved that the pressure was inversely proportional to the volume. It is to be noted that the agreement is not perfect. This was ascribed by Boyle to the difficulties of making perfectly accurate observations. In modern language, we would say that the two columns agreed within the limits of error of the experiment. Later and more exact experiments show that no real gas obeys Boyle's law exactly over large variations of pressure, but that observation in no way invalidates his conclusions.

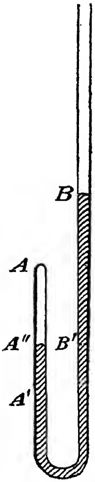


FIG. 59.—Boyle's apparatus for studying the relation between the pressure and volume of a gas. The volume AA'' decreases as the pressure exerted by the mercury column BB' is increased.

26. There is another way of expressing experimental results that was not used in Boyle's time but that has many advantages. If any two measurable quantities are related so that some particular measurement of one can be correlated with some particular measurement of the other, we can represent a series of measurements of these quantities by plotting them on a graph. In the experiment we are discussing, each number in the second column

measures a pressure corresponding to the volume measured by the corresponding number in the first column. They can therefore be plotted on a graph as in Fig. 60. The advantage of this procedure is that it does not depend on the knowledge of the mathematical relation between the two quantities. In fact it is

ORIGINAL DATA OF ROBERT BOYLE, 1680

Length of air column, in.	Length of Hg column + $29\frac{1}{8}$, in.	Calculated length of Hg, in.
36	$29\frac{2}{16}$	$29\frac{2}{16}$
$34\frac{1}{2}$	$30\frac{9}{16}$	$30\frac{6}{16}$
33	$31\frac{15}{16}$	$31\frac{12}{16}$
$31\frac{1}{2}$	$33\frac{8}{16}$	$33\frac{1}{7}$
30	$35\frac{5}{16}$	35
$28\frac{1}{2}$	37	$36\frac{15}{19}$
27	$39\frac{5}{16}$	$38\frac{7}{8}$
$25\frac{1}{2}$	$41\frac{10}{16}$	$41\frac{2}{17}$
24	$44\frac{3}{16}$	$43\frac{11}{16}$
$22\frac{1}{2}$	$47\frac{1}{16}$	$46\frac{3}{5}$
21	$50\frac{5}{16}$	50
$19\frac{1}{2}$	$54\frac{5}{16}$	$53\frac{10}{13}$
18	$58\frac{13}{16}$	$58\frac{2}{8}$
$17\frac{1}{4}$	$61\frac{5}{16}$	$60\frac{18}{23}$
$16\frac{1}{2}$	$64\frac{1}{16}$	$63\frac{11}{16}$
$15\frac{3}{4}$	$67\frac{1}{16}$	$66\frac{4}{7}$
15	$70\frac{11}{16}$	70
$14\frac{1}{4}$	$74\frac{2}{16}$	$73\frac{11}{19}$
$13\frac{1}{2}$	$77\frac{14}{16}$	$77\frac{2}{3}$
$12\frac{3}{4}$	$82\frac{12}{16}$	$82\frac{4}{17}$
12	$87\frac{4}{16}$	$87\frac{3}{8}$
$11\frac{1}{4}$	$93\frac{1}{16}$	$93\frac{1}{5}$
$10\frac{1}{2}$	$100\frac{7}{16}$	$99\frac{6}{7}$
$9\frac{3}{4}$	$107\frac{13}{16}$	$107\frac{7}{13}$
9	$117\frac{9}{16}$	$116\frac{4}{8}$

From R. Boyle, "A Defence of the Doctrine Touching the Spring and Weight of the Air," London, 1682, p. 58.

often very useful when there is no such relation. If there is such a relation, its nature is often suggested by the shape of the graph. Thus the student of coordinate geometry will guess at once that the points in Fig. 60 lie on a hyperbola given by the equation $PV =$ a constant. This is of course equivalent to the relation between P and V already stated, namely, that P is inversely pro-

portional to V and vice versa. We could plot the mathematical curve PV = a constant and see how closely the experimental points come to it. This would be the graphical equivalent of the calculations given by Boyle in the last column of the previous table. On the scale on which the points are plotted, they nearly

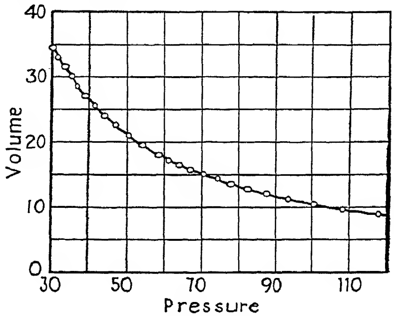


FIG. 60.—Graph of Boyle's original data. The points lie closely on a hyperbola, BV = a constant, showing that the pressure is inversely proportional to the volume.

all fall on such a curve. Our final statement of Boyle's law is as follows:

The volume occupied by a given amount of any permanent gas is inversely proportional to the pressure in the gas, provided the gas is kept at a constant temperature.

27. For the sake of precision in the wording of this law, three qualifications have been added to the bald statement. The first of these, that the amount of gas dealt with must remain unchanged, is obvious. The third requirement of constant temperature follows at once from the fact of thermal expansion discussed in the previous section. But what is meant by the second qualification, a "permanent" gas? The "permanent" gases, so christened at a time when no one had succeeded in liquefying them, include nitrogen, oxygen, hydrogen, and helium. They are found to conform very closely to Boyle's law over the ordinary range of temperatures and pressures. On the other hand, such gases as water vapor, carbon dioxide, and ammonia show considerable departures from Boyle's law except at very low pressures or high temperatures.

28. In mechanics it is often found helpful to talk of frictionless pulleys and perfectly smooth planes even though they can never

be realized in practice. Similarly, in studying gases, it is helpful to talk of an "ideal gas" which may be defined as one that would obey Boyle's law exactly at all temperatures and pressures. It would also obey Charles' law exactly at all temperatures and pressures. It is most nearly approximated by helium.

The Product of a Pressure and a Volume

29. For several paragraphs, we have been discussing the product of the pressure and the volume of a gas, and this product is going to keep recurring throughout this chapter and the next. This is a good point at which to consider what kind of a quantity such a product is. We can see quickly that it is an energy. This appears immediately from the units in which pressure and volume are expressed. In the m.k.s. system, volume is expressed in cubic meters and pressure in newtons per square meter. Written symbolically

$$\text{Pressure} = \text{newtons per square meter.}$$

and

$$\text{Volume} = \text{cubic meters;}$$

therefore, pressure $\times$ volume = newtons $\times$ cubic meters/square meters = newton-meters = joules. We find a somewhat more obviously physical approach to the same question if we consider the work done in compressing the gas in a cylinder such as is shown in Fig. 61. The work done is the total force on the piston times the distance the piston moves. But if the piston moves only so short a distance that the pressure can be considered constant, the total force on the piston is simply the pressure P, times A, the area of the piston, *i.e.*, PA. Hence, if the piston moves a distance d, the work is PAd; but the change in the volume of the gas which we may call $(V - V') = Ad$, therefore the work done in compressing the gas is $P(V - V')$, *i.e.*, the difference between two products of volume and pressure.

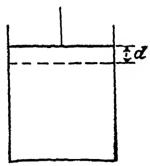


FIG. 61.—A cylinder with a movable piston.

Combination of Charles' Law and Boyle's Law

30. In Par. 24, we pointed out that we had to consider three variables in dealing with any given sample of gas, namely, the volume, the pressure, and the temperature. We saw that these were

three mutually dependent variables and that once the relations between any two pairs of them were established their mutual interrelation was completely determined. We have now studied the relation between volume and temperature as given by Charles' law and the relation between volume and pressure as given by Boyle's law. Therefore we are ready to set up a general relation that will give the value of any one of the three variables in terms of the other two. We effect this combination of the two previous results as follows:

31. Suppose we start with a certain quantity of gas at a temperature of 273°K. (0°C.), volume V_0, and pressure P_0. In other words, the initial state of the gas is characterized by the three coordinates 273, V_0, and P_0. Now keeping the pressure constant, raise the temperature to some higher value $T°K$. The volume will increase to some new value V' given by

$$V' = \frac{V_0 T}{273} \tag{9}$$

but the gas will still be at the pressure P_0. Now keep the temperature constant at the new value, and increase the pressure to some new value P. The volume will decrease to a smaller value V, but by Boyle's law the product of the volume and pressure remains constant, *i.e.*, $PV = P_0V'$. Substituting for V' from Eq. (9) above, we have

$$PV = \frac{P_0 V_0}{273} T \tag{10}$$

which is a perfectly general relation between the pressure, volume, and temperature for a given quantity of a perfect gas. In other words, for a given quantity of gas the product of the volume and the pressure is proportional to the absolute temperature.

32. Let us illustrate the preceding argument by a specific example. Suppose we start with 10 g. of air (density = 1.2 g./l.) at a pressure of 1 atm.; then $V_0 = 10/1.2 = 8.3$ l. Then P_0V_0 is 1×8.3 l.-atm., and the product of the pressure and volume for 10 g. of air at any temperature T will be given by

$$PV = \frac{P_0 V_0 T}{273} = T \times 1 \times 8.3 \text{ l.-atm.}$$

From this equation, we can find by substitution how many liters volume would be occupied by 10 g. of air at any particular

number of atmospheres pressure at any particular absolute temperature or we can find the pressure, given the temperature and volume, or the temperature, given the volume and pressure. If the pressures were measured in newtons per square meter and the volume in cubic meters, the constant of proportionality between PV and T would be different.

33. Suppose that instead of taking 10 g. of air as an example we take one mole (see Chap. I, Par. 37) of a pure gas, *e.g.*, 4 g. of helium. Then again, we have

$$PV = \frac{P_0V_0}{273}T;$$

and if we calculate the volume at 0°C. and one atmospheric pressure as before, we find that

$$V_0 = \frac{\text{mass}}{\text{density}} = \frac{4}{0.178} = 22.4 \text{ l.} = 0.0224 \text{ cu. m.}$$

If we take a mole of oxygen, *i.e.*, 32 g., we find

$$V_0 = \frac{32}{1.429} = 22.4 \text{ l.}$$

and, in general, we find that one mole of any gas at zero degrees and atmospheric pressure has a volume of 22.4 l. Consequently we can rewrite our gas law as follows:

$$PV = \frac{m}{M}RT \tag{11}$$

where m is the mass of the gas under consideration, M is its molecular weight, and R is a universal constant whose numerical value depends on the units used.

34. We wish to emphasize the general nature of Eq. (11). We might express the pressure in pounds per square inch, the volume in bushels, and the mass of the gas in ounces, but we would still be able to get a value of R that would make the preceding equation valid so long as the temperature was measured from the absolute zero. (Since the molecular weight is merely a ratio, it has always the same value in any system of units.) Obviously we are not likely to use the system of units just suggested. In the examples we have given, the pressure is measured in atmospheres, the volume in liters, and the mass in grams;

rather mixed units but sometimes the most convenient. In this system, the numerical value of R is evidently

$$R = \frac{22.4}{273}.$$

35. In the first chapter, we defined the gram molecule, or mole, as the quantity of a substance that has a weight in grams numerically equal to the molecular weight of the substance. This quantity is so firmly fixed in chemical usage that we did not venture to replace it by a kilogram molecule, defined in the same way in terms of the kilogram instead of the gram, although the kilogram molecule is the quantity of this type appropriate to the m.k.s. system of units. Evidently, if the mass of the gas is expressed in grams, the factor m/M in Eq. (11) is the number of moles or gram molecules of gas under consideration; and if the mass of the gas is expressed in kilograms, the factor m/M is the number of kilogram molecules under consideration. In either case, an appropriate value of R can be found for a given choice of units for the other quantities. These other quantities might be expressed in m.k.s. or c.g.s. units or in a mixed system such as is used in the examples of Pars. 32 and 33. We shall select three systems that give three different values of R as follows:

1. If the pressure is measured in atmospheres, the volume in liters, and the mass in grams, we have the system of our examples, one which is of common chemical usage, and find

$$R = \frac{22.4}{273} = 0.082 \frac{\text{l. atm.}}{\text{mole deg.}}$$

where it is to be remembered that a liter-atmosphere is an energy unit (= 101.3 joules).

2. If we use the m.k.s. system for everything but the mass, *i.e.*, the pressure is in newtons per square meter, the volume in cubic meters, and the mass in grams, and substitute joules for newton-meters, we have

$$R = 8.32 \frac{\text{joules}}{\text{mole deg.}}$$

3. If we go completely over to the m.k.s. system, expressing the mass in kilograms and the other quantities as in (2), we get

$$R = 8{,}320 \frac{\text{joules}}{\text{kg. molecule deg.}}$$

36. We have deduced the most general form of the gas law from the experimentally determined densities and the molecular weight. In practice, the molecular weights of gases can be determined by applying the gas law. Historically, the gas law was discovered as the result of Gay-Lussac's study of the laws of combining volumes in chemical reactions between gases and of his study of the thermal expansion of gases. Avogadro saw the implications of the results of Gay-Lussac. He stated that:

Equal volumes of all gases under the same conditions of temperature and pressure contain the same number of molecules.

37. That this does follow from the gas law is evident if we remember that the molecular weight is proportional to the weight of an individual molecule so that a gram molecule of one substance has the same number of molecules in it as a gram molecule of any other substance. Therefore, if the volume occupied by a gram molecule of a gas at NTP is the same for all gases, the number of molecules per cubic centimeter must be the same for all gases.

Quantity of Heat. Specific Heats

38. If heat flows into a body, its temperature rises. This sounds obvious enough but involves two new notions, quantity of heat and conduction of heat. We can get a rough intuitive notion of the relative amounts of heat required to produce different rises of temperature by observing the time of heating required. We know it takes more heat to cause a given rise in temperature for 10 kg. of water than for 1 kg. and more to raise 1 kg. 10° than 1°. We define the unit quantity of heat as the amount of heat required to raise 1 kg. of water from 15 to 16°C. and call this quantity the kilocalorie or, more specifically, the kilogram-calorie. For everything but the most precise work, the heat required to raise 1 kg. of water 1°C. at any temperature between freezing and boiling may be taken as 1 kg.-cal. One calorie or gram-calorie is one one-thousandth of a kilogram-calorie.

39. When other substances are compared with water, it is found that they have smaller heat capacity, *i.e.*, their temperature rises more quickly under similar circumstances. *The ratio of the amount of heat required to produce a given rise in temperature in a given mass of a substance to that required to produce an equal rise in an equal mass of water is called the specific heat of the sub-*

stance. Thus the specific heat of a substance is equal to the number of kilocalories necessary to raise the temperature of one kilogram of the substance one degree centigrade. The specific heats of a number of common substances are given below.

TABLE 7.—SPECIFIC HEATS

Water at 4°	1.0049	Copper	0.0931
Water at 15°	1.0000	Zinc	0.0935
Water at 30°	0.9971	Iron	0.114
Ice at 0°	0.502	Sulfur	0.176
Lead	0.0310	Aluminum	0.217
Mercury	0.0331	Lithium	0.941
Tin	0.0562	Crown glass	0.16
Silver	0.0570	Flint glass	0.117
		Thermometer glass	0.199

40. From Eq. (10), it is evident that a rise in the temperature of a gas corresponds to an increase in the product PV, which in turn according to Par. 29 corresponds to an increase in energy. Therefore the flow of heat causing a rise in temperature is at the same time a flow of energy causing an increase in energy content. Hence heat is a form of energy, a conclusion which we shall confirm presently by a different argument. If heat is a form of energy, it must obey the law of conservation of energy. That is, if a number of different substances at different temperatures are mixed in a container which is heat insulated, the heat lost by the hotter substances must just equal that gained by the colder.

41. A container used for this purpose is so constructed that the amount of heat that flows to it or from it during the course of the experiment is negligible. It is called a calorimeter. Suppose a calorimeter of mass m_1, specific heat s_1, and temperature t_1 has poured into it m_2 kg. of liquid of specific heat s_2, at temperature t_2. A mass m_3 of a solid of specific heat s_3 and temperature t_3 is then dropped into the vessel and the combination allowed to stand. Heat will flow from the warmer to the colder parts of the system until they are finally all at the same temperature. Call this final temperature t, and apply the principle of conservation of energy. Expressed mathematically, the result is

$$m_1s_1(t - t_1) + m_2s_2(t - t_2) + m_3s_3(t - t_3) = 0 \qquad (12)$$

or, in words, the net change in the heat energy present is zero. The three different terms give the gain or loss of heat of each

of the three substances, a gain if the temperature change is positive, a loss if it is negative. Obviously at least one term must be positive and one negative. Further, each term is proportional to the mass and specific heat of the substance in question. This follows from our definitions of specific heat and quantity of heat. The equation has been written for three substances but may be extended to include more by adding appropriate terms. If the temperatures and masses in such an experiment are measured and the specific heat of two of the substances known, that of the third may be determined. In the common "method of mixtures" for determining specific heats, the calorimeter is copper and the liquid water, the specific heat of the solid being the unknown quantity.

Conduction of Heat

42. Wrapping a thermometer bulb with asbestos would not improve its quality as a temperature-measuring device. A thermometer works because it gradually takes up the temperature of its surroundings. Heat flows from it or to it until it comes to that temperature. The same thing happens with any bodies brought into contact with each other after being initially at different temperatures. The rate at which heat flows from one body to another or from one part of a body to another depends on the material through which it flows. Asbestos is a bad conductor of heat. Consequently an asbestos-wrapped thermometer would be very slow in responding to changes of temperature in its surroundings. Metals, on the other hand, are good conductors of heat just as they are good conductors of electric current. But in practical physics it is more often important to discourage the flow of heat than to encourage it. Large differences of temperature are often of great usefulness though the whole tendency of nature is to eliminate them. Therefore in ordinary life, one is more apt to hear of heat insulators than of heat conductors. The very worst conductor known, therefore the best insulator, is nothing at all. That is the reason for evacuating the space between the inner and outer walls of a thermos bottle.

Mechanical Equivalent of Heat

43. That heat is a form of energy is a familiar fact, for we know that by doing mechanical work we may arrange to have

mechanical energy used up with the appearance of heat. Conversely, it is well-known that heat may be converted into mechanical energy as is done in steam engines or internal-combustion engines. The branch of physics that is called *thermodynamics*, as its name implies, is concerned with the detailed study of the processes whereby energy is transformed from the mechanical forms, like potential and kinetic energy, into the heat forms, or vice versa.

44. The question now comes up: If quantity of heat is really synonymous with quantity of energy of a particular form called heat energy, how much energy is there, measured in joules, in 1 kg.-cal. of heat? In other words, how much mechanical energy, in joules, must be converted into heat energy in order to raise the temperature of 1 kg. of water through 1°C.? This quantity, which is the factor of conversion between heat energy measured in kilocalorie units and heat energy measured in one of the mechanical energy units with which we are already familiar, is known as the *mechanical equivalent of heat*.

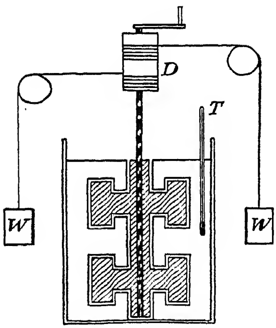


FIG. 62.—Apparatus for determining the mechanical equivalent of heat. The weights W fall, turning the paddle wheel by means of the drum D. The thermometer T measures the consequent rise in temperature of the water in the calorimeter.

45. There are a number of ways in which the mechanical equivalent of heat can be measured. The most direct way and the way in which it was first accurately measured depends on fluid friction. A known amount of a liquid, usually water, is put in a calorimeter. Then some sort of paddle wheel is put in the liquid in such a way that the amount of energy used to make it turn can be easily measured. This can be accomplished by connecting the paddle wheel through pulleys to a falling weight. The potential energy of the weight is converted to kinetic energy of the paddle wheel and the liquid it churns up. The viscosity of the liquid finally turns this kinetic energy into heat. The consequent rise in temperature of the liquid is observed, and the heat energy absorbed is computed. This is then compared with the potential energy that has been lost by the falling weight and the mechanical

equivalent of heat calculated. The difficulty of getting a calorimeter that is so perfectly insulated that it does not exchange heat with its surroundings, the frictional losses of energy in the mechanical drive of the paddle wheel, and other troubles make this experiment hard to perform with great accuracy.

46. In Fig. 62, we have a cross-sectional view of an apparatus of this kind similar to one used originally by Joule in 1845. The outcome of many careful experiments of this type is that the amount of energy corresponding to the kilocalorie is known to be 4,185 joules, or approximately 4.2×10^3 joules.

47. Another method of finding how much energy it takes to warm up 1 kg. of water through 1°C. is by electrical heating. When an electric current passes through a wire, electrical energy is converted into heat. This fact is, of course, the basis of a great many domestic uses of electric energy in heaters, toasters, stoves, and so on. If, therefore, we put a coil of resistance wire in a calorimeter, the electrical energy consumed will appear as heat energy which goes to warm up the water in the calorimeter. After we have learned the methods of measuring electrical energy, we shall see the details of the way in which the energy needed to heat up a given amount of water may be measured electrically. Such a method is more easily developed into a precision method of measuring the mechanical equivalent of heat than the mechanical methods, although it is not so direct from the standpoint of the beginning student of physics. Both types of experiment give the same value to within the inevitable experimental errors.

48. The mechanical equivalent of heat is really quite a "large" quantity from the standpoint of everyday quantities with which we are familiar. A kilogram of water is not very much (a drinking glass full is about a quarter of a kilogram) and 1°C. is not a very great rise in temperature (from freezing up to a comfortable room temperature is 20°). Yet the amount of energy needed to produce the 1° rise in only 1 kg. is 4,185 joules. This is roughly as much as is needed to lift a weight of 40 kg. through a distance of 10 m.

Chemical Energy and Heat

49. In connection with our remarks on the work done by human beings, we have already had occasion to mention still another form of energy which we loosely called *chemical energy.*

Let us now turn to a consideration of the way in which this is measured. Without attempting to give a really precise definition of what we mean by chemical energy—for that is quite difficult—we may say that we mean the energy changes accompanying a chemical reaction. Most familiar to us in everyday life are the chemical reactions that involve combustion of fuels with a consequent generation of heat, which we have already recognized as a form of energy.

50. That we really have to do here with a transformation of energy from one form to another is shown by experiments in chemistry in which conditions are arranged so that the chemical reaction is made to go in the other direction from that occurring in combustion. To cause this reverse reaction, heat must be supplied and this heat is transformed into chemical energy by the reaction; thus heat disappears in an amount equal to that which would appear in the corresponding combustion.

51. For simplicity let us consider the simplest fuel we can think of, hydrogen gas. In chemistry we learn that the burning of hydrogen means a reaction of hydrogen gas with oxygen gas in which water vapor is produced and heat is given off according to the reaction

$$2H_2 + O_2 \rightarrow 2H_2O + \text{heat energy.}$$

We here follow the convention made earlier and suppose that the equation so written refers to definite quantities of the gas materials involved, namely, to gram-molecular weights or moles. In this case, as the molecular weight of hydrogen is 2 and that of oxygen 32, the preceding equation refers specifically to the burning of 4 g. of hydrogen gas, which requires the supply of 32 g. of oxygen gas with a production of 36 g. of water vapor.

52. Just as we speak of a moving body as possessing kinetic energy because of its capability of doing work against a retarding force which brings it to rest, so here we speak of the hydrogen and oxygen as possessing "chemical energy" because they have the capability of supplying energy in another previously recognized form when they react chemically. Also just as we took the amount of the work that the moving body could do in being brought to rest as a measure of its kinetic energy, so here we take the amount of heat energy that is produced by the reaction as a measure of the chemical energy of the hydrogen and oxygen gas.

Or more precisely, it is really the *change* in chemical energy, the difference in chemical energy between hydrogen and oxygen in the form of separate elements and the chemical energy of the same hydrogen and oxygen in the form of water vapor.

53. To be still more precise, it is necessary to state the condition of the hydrogen gas and oxygen gas, as to temperature and pressure before the reaction, and the condition of the water vapor produced as a result of the reaction. These refinements are in no way essential to a grasp of the main idea, although obviously they are very important if we wish to make accurate statements about the measurement of chemical energy. We do not discuss them more fully here for fear of obscuring the main point.

54. The amount of heat energy evolved in a chemical reaction is measured by arranging to have the reaction take place in a calorimeter. In this way, the heat evolved in many reactions has been studied by physical chemists. In particular, it is found that in the reaction of hydrogen and oxygen gas under standard conditions to form water vapor first and then by subsequent cooling to room temperature to produce finally liquid water, the amount of heat liberated is about 69 kg.-cal./g.-mole, *i.e.*, per 18 g. of water formed. This is about 3.8 kg.-cal./g. of water formed, in other words about thirty-eight times as much heat as is required to warm up the same amount of water from the freezing point to the boiling point.

55. It is rather difficult to arrange matters so that the reaction goes in the reverse direction, *i.e.*, so that water vapor is, by application of heat, dissociated back into hydrogen gas and oxygen gas. But it can be done, and it is found that the amount of heat absorbed in the process is the same as that given out in the formation of water. Therefore there is a transformation of heat back into chemical energy in the process. Thus chemical energy may be transformed into heat energy and heat energy into chemical energy under proper conditions in a manner similar to the way in which potential energy may be converted into kinetic energy, and vice versa, in simple mechanical apparatus.

56. The most familiar way of reversing this chemical reaction is by means of electrolysis. We arrange a cell of acidulated water (acidulated simply to make the material a better conductor of electricity) and pass a current through it by dipping into it two metal electrodes which are connected to a source of electric

power. At one electrode, bubbles of hydrogen gas appear, at the other bubbles of oxygen gas appear. It is an old and familiar experiment. Moreover it is the basis of the industrial method of preparing hydrogen gas for use in oxyhydrogen welding where the heat of the reaction we are considering is used in welding structural steel. If we measure the electrical energy necessary to dissociate water into hydrogen and oxygen in this way, we find that it is not simply the amount of energy necessary to drive a current through the cell against the electrical resistance of the cell. That may be made quite small, by proper design of the cell, relative to the whole amount of electrical energy used. Most of the electrical energy goes into the actual work of pulling apart the water molecules to form hydrogen and oxygen gas, just the reverse of the reaction that occurs when hydrogen and oxygen gas burn to form water. And the amount of energy needed to do this is accurately the same as the amount that is set free as heat, when the gases burn. In other words there is, in the cell, a conversion of electrical energy into chemical energy which may later be obtained as heat energy when the gases are allowed to recombine by burning in the oxyhydrogen welder's torch.

SUMMARY

Change of temperature is measured by the change in some property of a substance as it gets hotter or colder. The property most frequently used is the volume. The temperatures of certain easily standardized conditions are arbitrarily taken as reference points. The legal standard for measuring temperature is a hydrogen-gas thermometer. To a close approximation, the volume of any gas at a temperature of t°C. is given by

$$V_t = V_0\left(1 + \frac{t}{273}\right)$$

where V_0 is the volume at 0°C., the melting point of ice. This is the law of Charles and Gay Lussac. According to this law, there is a certain temperature -273°C. called the absolute zero, where the volume of a "perfect gas" would be zero. Temperatures measured from this point are called absolute temperatures. This temperature scale can also be established by an abstract argument. All substances change their volume with the temperature according to a law whose exact form depends on the particular substance.

In 1660, Boyle showed that the product of pressure and volume was constant for a given quantity of gas at a constant temperature. The product of pressure and volume is energy. Boyle's law and Charles' law can be combined to give the following general gas law:

$$PV = \frac{P_0V_0}{273}T$$

where T is the absolute temperature. Because of the experimental result that one mole of any gas occupies a volume of 22.4 l. at NTP, this gas law may be written

$$PV = \frac{m}{M}RT$$

where m is the mass of the gas, M its molecular weight, and R a universal constant.

The unit of quantity of heat is the kilocalorie, defined as the heat required to raise the temperature of 1 kg. of water 1°C. The specific heat of a substance is equal to the number of kilocalories needed to raise the temperature of 1 kg. of the substance 1°C. Heat is a form of energy, and 1 kg.-cal. is found experimentally equivalent to 4,185 joules of mechanical energy. Heat is absorbed or given off in chemical reactions.

ILLUSTRATIVE PROBLEMS

1. Jena thermometer glass has a coefficient of cubical expansion of 1.272×10^{-5} per degree centigrade. The coefficient of cubical expansion of mercury is 1.8182×10^{-4} per degree centigrade. How much error at a temperature of 100°C. would the expansion of the glass make if it were not allowed for in the calibration of a mercury thermometer?

The distance between divisions on the thermometer scale is determined by the relative expansion of glass and mercury. The volume per degree between graduations is the increase in volume of mercury minus the increase in volume of glass for 1° rise in temperature. The increase in volume for 1° rise may be found from Eq. (7), the terms after αt being neglected. Thus

$$V_t = V_0(1 + \alpha t) \qquad \text{and} \qquad V_t - V_0 = \Delta V = V_0\ \alpha t.$$

The volume per degree of the thermometer stem is then

$$\Delta V_{\text{Hg}} - \Delta V_{\text{gl}} = V_0(\alpha_{\text{Hg}} - \alpha_{\text{gl}}) \times 1\,°\text{C}.$$

The error in the temperature reading contributed by neglecting the volume expansion of the glass will be the increase in volume, due to the expansion of the glass, divided by the volume per degree of the thermometer stem.

$$\frac{V_0\,\alpha_{gl}\,100°\text{C.}}{V_0(\alpha_{Hg} - \alpha_{gl})} = \frac{1.272 \times 10^{-5} \times 1/°\text{C.} \times 100°\text{C.}}{18.182 \times 10^{-5} \times 1/°\text{C.} - 1.272 \times 10^{-5} \times 1/°\text{C.}}$$
$$= 7.5°\text{C.}$$

2. The temperature of a steel rail may vary from −40°C. (−40°F.) to 60°C. (140°F.). If the length of the rail at 0°C. is 50 ft., what is the difference in length of the rail at the two extreme temperatures? If the length at 60°C. were computed from the length at −40°C. directly, what would be the percentage of error in your answer?

The change in length with temperature is given by Eq. (8).

$$L_t = L_0(1 + \alpha t).$$

We obtain the lengths at −40 and 60°C. by direct substitution in this formula. The coefficient of expansion of steel from Table 6, page 145, is 1.332×10^{-5} per degree centigrade.

$$L_{-40} = 50 \text{ ft.} [1 + 1.332 \times 10^{-5} \times 1/°\text{C.} (-40°\text{C.})]$$
$$= 50 \text{ ft.} - 0.02664 \text{ ft.} = 49.97336 \text{ ft.}$$
$$L_{60} = 50 \text{ ft.} [1 + 1.332 \times 10^{-5} \times 1/°\text{C.} \times 60°\text{C.}]$$
$$= 50 \text{ ft.} + 0.03996 \text{ ft.} = 50.03996 \text{ ft.}$$

Therefore the difference in length of the rail at the two extreme temperatures is

$$L_{60} - L_{-40} = 50.03996 - 49.97336 = 0.0666 \text{ ft.} = 0.80 \text{ in.}$$

Starting with a 49.97336-ft. rail at −40°C. and calculating the length at 60°C. directly from Eq. (8) gives

$$L_{60} = 49.97336\ (1 + 1.332 \times 10^{-5} \times 1/°\text{C.} \times 100°\text{C.})$$
$$= 49.97336 \text{ ft.} + 0.06656 \text{ ft.} = 50.03992 \text{ ft.}$$

An error of 50.03996 ft. − 50.03994 ft. = 0.00004 ft., or

$$\frac{0.00004 \text{ ft.} \times 100}{50.03996 \text{ ft.}} = 8 \times 10^{-5}\,\%.$$

3. A bubble of marsh gas is formed at the bottom of a lake 20 ft. deep. The temperature of the water at this depth is 10°C. The atmospheric pressure is equal to that caused by a column of mercury 74 cm. high. The gas is caught at the surface and occupies a volume of 8.4 cm.³ at a temperature of 18°C. What volume did it occupy at the bottom of the lake? This problem is an example of the law of perfect gases $PV = \frac{P_0V_0}{273}T$. This law must be applied to a certain mass of gas contained in the bubble first at the surface of the lake and then at the bottom of the lake. If we write the quantities at the surface of the lake with a subscript 1 and those at the bottom with a subscript 2, then

$$\text{At the top } P_1V_1 = \frac{P_0V_0}{273}T_1$$

$$\text{At the bottom } P_2V_2 = \frac{P_0V_0}{273}T_2$$

whence

$$\frac{P_1V_1}{T_1} = \frac{P_2V_2}{T_2} = \frac{P_0V_0}{273}. \tag{13}$$

The numerical values of these quantities are
At the top

$$P_1 = 74 \text{ cm. Hg}$$
$$V_1 = 8.4 \text{ cm.}^3$$
$$T_1 = 273°\text{C.} + 18°\text{C.} = 291°\text{K.}$$

At the bottom

$$P_2 = 74 \text{ cm.} + \frac{20 \text{ ft.} \times 2.54 \text{ cm./in.} \times 12 \text{ in./ft.} \times 1 \text{ g./cm.}^3}{13.6 \text{ g./cm.}^3}$$
$$= 74 \text{ cm.} + 44.8 \text{ cm.} = 118.8 \text{ cm.}$$
$$V_2 = ?$$
$$T_2 = 273°\text{C.} + 10°\text{C.} = 283°\text{K.}$$

Inserting these in Eq. (13), we obtain

$$\frac{74 \text{ cm.} \times 8.4 \text{ cm.}^3}{291°\text{K.}} = \frac{118.8 \text{ cm.} \times V_2}{283°\text{K.}}$$
$$V_2 = \frac{74 \text{ cm.}}{118.8 \text{ cm.}} \times \frac{283°\text{K.}}{291°\text{K.}} \times 8.4 \text{ cm.}^3$$
$$= 5.1 \text{ cm.}^3$$

PROBLEMS

1. Where does alcohol freeze and boil on the Fahrenheit and centigrade scales?

2. What is the normal temperature of the human body on the Fahrenheit and centigrade scales?

3. If the electrical resistance of a platinum wire is 10 at 0°C. and 13.92 at 100°C. and varies linearly with the temperature in between, write an expression for the temperature in terms of the resistance to serve as a resistance thermometer.

4. A 1-l. pyrex flask is full of mercury at 0°C. If the coefficient of cubical expansion of pyrex is 1.08×10^{-5} per degree centigrade, how much mercury will spill out if the flask and contents are warmed 50°C.?

5. The volume of a thermometer bulb is 0.80 cc. and it is filled with mercury. If the mercury column in the capillary tube attached to the bulb rises 1 mm. for a 1°C. rise in temperature, what is the cross section of the capillary? (Neglect the expansion of the glass and of the mercury in the capillary.)

6. A thermometer is desired for demonstration purposes that will give a rise of 1 cm., for a 1°C. rise in temperature, in a tube of 5 sq. mm. cross section when it is filled with alcohol. How large a volume of alcohol must the bulb of the thermometer hold?

7. Two identical thermometers like that in Prob. 6 are filled with alcohol and calibrated with uniform graduations marking degrees from −10 to

100°C. One of them is then emptied and refilled with water. The two bulbs are then placed in an oil bath whose temperature is gradually raised from 0 to 50°C. Plot a curve showing the readings of the water-filled thermometer plotted against those of the alcohol-filled thermometer from 0 to 50°C. Calculate the reading of the water thermometer for every degree rise of the alcohol thermometer from 1 to 10 and then for every five degrees from 10 to 50, and plot a curve showing one as a function of the other. (Calculations to tenths of a degree will be sufficiently accurate.)

8. How much does the length of an iron rail 40 ft. long at 0°C. change by reason of a change of temperature from −20 to 35°C?

9. The engine temperature in a modern automobile is measured by a metal thermal-expansion thermometer. If this thermometer is 5 cm. long and has a thermal expansion of 30×10^{-6} per degree centigrade, about how many times must the motion of the free end of this thermometric unit be magnified to give the readings shown on the instrument panel of the car?

10. A steel wire of 1 sq. mm. cross section is drawn taut between fixed supports with a tension of 1 kg. weight at a temperature of 15°C. The temperature then falls to 0°C. Assuming that the wire is not stretched beyond its elastic limit, what is the tension in it?

11. A steel wire 10 m. long and 0.1 sq. cm. cross section is alternately stretched and released by a force of 100 newtons (about 10 kg. weight). If this occurs one hundred times and Hooke's law holds throughout, how much heat is generated in the wire?

12. A steel I beam of 10 sq. cm. cross section is warmed 50°C. Find the force of compression necessary to prevent expansion.

13. A brass plug is to be shrunk into a hole in a steel plate. If at 20°C. the plug is 0.01 mm. too large in diameter for a 2-cm. hole, at what temperature will it fit?

14. Calculate the gas constant in joules per gram per degree for oxygen.

15. What will be the resultant temperature if 500 g. of water at 90° are poured into a copper calorimeter weighing 200 g., and containing 500 g. of water at 20°?

16. A block of silver weighing 1,000 g. at 95°C. is immersed in 500 g. of water at 20°C. contained in a silver calorimeter weighing 200 g. The resultant temperature is 27.4°C. What is the specific heat of silver?

17. If the pressure of a gas in a fixed volume is 76 cm. of mercury at 15°C. and then by reason of a rise of temperature becomes 114 cm. of mercury, what is the new temperature?

18. If the temperature of a gas is raised from 27 to 100°C. and the pressure is doubled, what is the ratio of the resulting volume to the original volume?

19. In Rumford's experiment on the heat developed in boring brass cannon, the amount of heat developed in heating the metal and the water was estimated by him as sufficient to raise the temperature of 26.58 lb. of water 180°F. The time of the experiment was 2 hr. 30 min. The work was actually done by two horses, but Rumford says that it could easily have been done by one horse. Calculate from these data a value of the mechanical equivalent of heat, assuming that $\frac{2}{3}$ hp. was expended in producing heat.

20. Illuminating gas costs 50 cts. per 1,000 cu. ft. Each cubic foot will give 1.3×10^5 g.-cal. of heat upon combustion. What will it cost to heat a 60-gal. (1 gal. = 3.8 l.) tank of water from 12 to 70°C., assuming an efficiency of 75 per cent? What will it cost to heat the same tank by electricity at $2\frac{1}{2}$ cts./kw.-hr. with 100 per cent efficiency?

21. What volume will be occupied by 100 g. of carbon monoxide at a pressure of 100 cm. of mercury and a temperature of 100°C.? What volume would be occupied by 100 g. of carbon dioxide under the same conditions?

22. What would be the numerical value of the coefficient of expansion of a gas if the original volume were measured at 100 instead of 0°C.?

23. What would be the value of the coefficient of cubical expansion of a gas at constant pressure if the Fahrenheit degree were used instead of the centigrade degree?

24. What would be the percentage change in the volume of a soap bubble if it were originally filled with air at 40°C. and then cooled to 20°C.?

25. Air at a temperature of 60°F. is pumped into an automobile tire until the gauge pressure is 35 lb./sq. in. After a trip in the hot sun, the temperature in the tire is 110°F. What is the pressure within the tire, if it is assumed that the volume change is negligible?

26. A diver releases a bubble of air which doubles in volume in rising to the surface of the water. At what depth is the diver?

27. Express the law of Charles and Gay-Lussac in terms of degrees Fahrenheit and the Fahrenheit zero. Convert the temperature in Prob. 17 to Fahrenheit, and use the new form of the gas law to solve the problem.

28. A lead bullet with a speed of 500 m./sec. strikes a wall and falls to the ground. Assuming that one-fourth of the energy goes into heating the bullet, what is its rise in temperature?

CHAPTER VIII

THE KINETIC THEORY OF GASES

1. In the first two chapters, we discussed the structure of matter in terms of atoms and molecules in a qualitative way. Then we spent some time learning how to deal with motion, force, and energy in a quantitative way. Finally, in the last chapter, the subject of heat was discussed and it was shown that heat is a form of energy. In the present chapter, we wish to show that heat is, in the words of Count Rumford, a "mode of motion." In particular, we want to show that the phenomena described in the last chapter can be explained by applying the laws of mechanics to the general assumptions of the atomic theory of matter as given in the first two chapters. In some cases, the explanation will be only qualitative but in other cases, such as the relation between the pressure and temperature of a gas, the explanation will be quantitative. Furthermore we shall see that in explaining these phenomena other theoretical relations will suggest themselves which can be tested by comparison with experimental results.

2. First of all, let us attempt to explain the pressure of a gas in terms of molecular motion. Our assumptions are that a gas is made up of a large number of very small particles in violent chaotic motion. We call these particles molecules. For the moment, we assume that the forces of attraction between the molecules may be neglected. That is, we assume that the gas is at so low a pressure that the average distance between molecules is large. We further assume that the molecules behave like perfectly hard elastic spheres so that when they collide with each other or with the walls of a containing vessel momentum and kinetic energy are both conserved. Granted these assumptions, let us see what effect is to be expected from a number of molecules enclosed in a container.

3. Let the container be a rectangular box of sides a, b, and c and therefore of volume $V = abc$, and choose axes of coordinates

parallel to the edges of this box as shown in Fig. 63. Let the number of molecules in the box be n and the mass of each molecule be μ. Let the walls of the container be perfectly smooth, perfectly elastic, and of infinite mass. If this is the case, a collision of a molecule with a wall will affect only the component of the molecule's velocity perpendicular to that wall and will leave the other two components of its velocity unchanged. Thus in Fig. 63 suppose a molecule has components of velocity u, v, and w, parallel to the x-, y-, and z-axes, respectively, and suppose it is all alone in the box so that its only collisions are with the walls; then the time it will take to travel from the wall $OADC$ to the wall $BGEF$ will be b/v no matter what values u and w have nor how many collisions the molecule makes with the other four walls during its passage from $OADC$ to $BGEF$. Similarly, the time of passage from the wall $OAGB$ to the wall $CDEF$ will be c/w and the time of passage from the wall $OBFC$ to the wall $AGED$ will be a/u. The three motions can therefore be treated quite independently. Let us deal with the first. Every time a molecule hits the wall $BGEF$, it bounces back elastically, *i.e.*, its component of momentum along the y-axis changes from μv to $-\mu v$, a total change of $2\mu v$. Since the wall is perfectly smooth, its other two components of momentum are unchanged. Since each molecule takes a time b/v to go from $OADC$ to $BGEF$, its time between collisions with $BGEF$ will be twice this and it will hit $BGEF$ $v/2b$ times every second. Therefore the total change in its momentum in 1 sec. will be $v/2b \times 2\mu v = \mu v^2/b$. If many molecules are hitting the wall all the time and reversing their momentum, there must be a continual force on the wall. It is this force which we wish to calculate. Suppose that in the box we have not only one molecule but n molecules having y-components of velocity equal to $v_1, v_2, \ldots v_n$, respectively. (We are assuming that n is still so small that the collision of the molecules with each other can be neglected.) Then the number of molecules hitting $BGEF$ every second is $(v_1 + v_2 + \cdots v_n)/2b$ and each one has its momentum

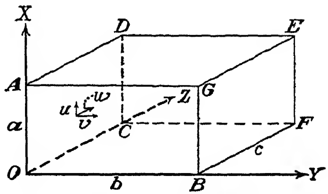


FIG. 63.—A rectangular box with its sides a, b, c parallel to the rectangular coordinate axes x, y, z. The velocities of a molecule parallel to the x-, y-, z- axes are u, v, w.

changed by $2\mu v$ at each impact. The total change in momentum divided by the time during which the change occurs must equal the average force exerted on the wall during that time. Therefore the average force on the wall $BGEF$ is the total change in momentum of the molecules hitting in one second, *i.e.*,

$$\frac{\mu v_1^2}{b} + \frac{\mu v_2^2}{b} + \cdots \frac{\mu v_n^2}{b} = \mu\frac{(v_1^2 + v_2^2 + \cdots v_n^2)}{b}. \tag{1}$$

If we let $\overline{v^2}$ be the average value of v^2, this reduces to $n\mu\overline{v^2}/b$. What we are interested in is not total force but force per unit area or pressure. The area of $BGEF$ is ac; therefore the pressure on $BGEF$ is

$$\frac{n\mu\overline{v^2}}{abc}. \tag{2}$$

The pressure on $AODC$ is obviously the same. By exactly analogous arguments, the pressures on the two sides $OAGB$ and $CDEF$ and on the two sides $OCFB$ and $ADEG$ are found to be

$$\frac{n\mu\overline{w^2}}{abc} \quad \text{and} \quad \frac{n\mu\overline{u^2}}{abc}, \text{ respectively,}$$

4. Our original assumption was that the motion of the molecules in a gas was perfectly random, that a molecule was just as likely to be moving one way as another. This means that the average magnitude of the velocity along one axis is the same as that along any other. Consequently, $\overline{u^2} = \overline{v^2} = \overline{w^2}$. Let $\overline{W^2}$ be the average of the squares of the resultant velocities of the molecules. Then, in Fig. 64, OAB is a right triangle in the horizontal yz-plane and OBC is a right triangle in a vertical plane. In OAB, the square of the hypotenuse OB equals the sum of the squares of the sides, *i.e.* $\overline{v^2} + \overline{w^2} = (OB)^2$. Similarly in OBC, $(OC)^2 = \overline{W^2} = (OB)^2 + (BC)^2 = \overline{u^2} + \overline{v^2} + \overline{w^2}$. Consequently

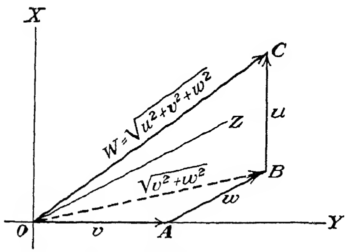


FIG. 64.—Addition of velocities by the Pythagorean theorem. $\overline{OB}^2 = \overline{v^2} + \overline{w^2}$. $\overline{W^2} = \overline{OB}^2 + \overline{u^2} = \overline{u^2} + \overline{v^2} + \overline{w^2}$.

$$\overline{W^2} = \overline{u^2} + \overline{v^2} + \overline{w^2} \qquad \text{or} \qquad \overline{u^2} = \overline{v^2} = \overline{w^2} = \tfrac{1}{3}\overline{W^2}. \tag{3}$$

So that the pressure on all the sides of the box is seen to be equal and to reduce to

$$P = \frac{n\mu\overline{W^2}}{3V} \tag{4}$$

where V is the volume of the box.

5. Before going further, let us apply these results to a particular case and see if they make sense. The product $n\mu$ is simply the total mass of the gas in the box and since it is divided by the volume V of the box we can rewrite Eq. (4) as

$$P = \overline{W^2} \tag{5}$$

where $\rho = n\mu/V$ is the density of the gas. Apply this equation to helium at 0°C. and atmospheric pressure. The density of helium is 0.178 kg./cu. m. under these conditions, and a pressure of 1 atm. equals 1.013×10^5 newtons/sq. m. Putting these values in Eq. (5), we have $1.013 \times 10^5 = 0.178\overline{W^2}/3$. Solving this equation for W, we find that $W = 13.1 \times 10^2$ m./sec. But in Chap. II we discussed an arrangement for measuring directly the velocity of molecules. The values determined there were of the order of 200 to 700 m./sec. for molecules of bismuth, and we see that the value just calculated for helium is of the same order of magnitude. We shall see presently that the higher value results from the lighter molecule. At least our theoretical deductions are all right so far. (It may be of interest to point out that molecular velocities were calculated in the foregoing manner many years before they were determined directly.)

6. Now let us compare our theoretical results with previous experimental ones in a more general way. The general law for the relation between the pressure, volume, and temperature of a gas was found to be $PV = mRT/M$; or if we express it in terms of the masses of the individual molecules and their numbers,

$$PV = \frac{n\mu}{N\mu}RT \tag{6}$$

where n is the number of molecules in the sample of mass m g. and N is Avogadro's number, the number of molecules in a mole, P and V are in m.k.s. units, and R has the appropriate value given in Par. 35 (2) of the last chapter. But our theoretical expression

is $PV = \frac{n\mu\overline{W^2}}{3}$. Therefore our theoretical and empirical expressions agree if

$$\frac{n\mu}{N\mu}RT = \frac{n\mu\overline{W^2}}{3} \qquad \text{or} \qquad \frac{1}{2}\mu\overline{W^2} = \frac{3}{2}\frac{R}{N}T. \tag{7}$$

Our kinetic theory of gases thus gives us the known law for the relation between pressure, volume, and absolute temperature, if we assume that each molecule has on an average a kinetic energy of $\frac{3}{2}\frac{R}{N}T$. We simplify this by writing $R/N = k$, a constant known as Boltzmann's constant, so that each molecule has a kinetic energy of $\frac{3}{2}kT$.

7. It was pointed out in Chap. VII, Par. 35, that the value of R depended on the units used for P and V. Since we naturally express velocities in meters per second, momenta in kilogram-meters per second, kinetic energies in joules, and forces in newtons when applying our kinetic theory, we want R in absolute units, *i.e.*, the volume in cubic meters and the pressure in newtons per square meter. This we have seen to be approximately

$$R = 8.3 \frac{\text{joules}}{\text{mole °C.}}$$

If we divide this value of R by the value of N, Avogadro's constant, we get

$$k = \frac{R}{N} = 1.38 \times 10^{-23} \frac{\text{joule}}{\text{molecule °C.}} \tag{8}$$

In other words, the average kinetic energy of a molecule of any gas at the absolute temperature T is $\frac{3}{2} \times 1.38 \times 10^{-23} \times T$ joules. We see now why the velocity of the helium molecules that we just calculated came out considerably higher than the measured velocities of the bismuth molecules given in Fig. 9, page 35. The kinetic energies of the bismuth molecules were higher since they emerged from a furnace at a temperature of several hundred degrees but their velocities were lower because of their much greater mass. Bismuth has an atomic weight of 209, and probably many of the molecules in this particular experiment were diatomic.

8. As one example of molecular velocity, let us calculate that of a nitrogen molecule at a temperature of 0°C. We have

$$\mu = \frac{M}{N} = \frac{28}{6.03 \times 10^{23}} = 4.64 \times 10^{-23}\text{ g.} = 4.64 \times 10^{-26}\text{ kg.}$$

Therefore

$$\frac{1}{2}\mu\overline{W^2} = \frac{3}{2}kT \qquad \text{and} \qquad \overline{W^2} = \frac{3kT}{\mu}$$

which gives, if $T = 273°$ abs. (0°C.),

$$\sqrt{\overline{W^2}} = \sqrt{\frac{3 \times 1.38 \times 10^{-23} \times 273}{4.64 \times 10^{-26}}} = \sqrt{2.43 \times 10^5}$$
$$= 4.94 \times 10^2 \text{ m./sec.}$$

We must point out that $\sqrt{\overline{W^2}} = W$ is not the average speed but the square root of the average of the squares of the various velocities, by no means the same. It is called the "root mean square" or r.m.s. speed.

9. We cannot say yet that we have "proved" that the kinetic theory is right. But we have seen that it agrees with the general empirical gas law if we assume that each molecule has on the average a kinetic energy of $\frac{3}{2}kT$ joules. Let us see whether this assumption can be tested in other ways. The direct measurement of molecular velocities we have already mentioned. The motion of molecules can be demonstrated by experiments on the diffusion of gases, and incidentally we demonstrated that hydrogen diffused more rapidly than air at the same temperature. According to our present result, the r.m.s. speeds of molecules vary directly as the square root of the absolute temperature and inversely as the square root of their masses or, in mathematical language, $W \propto \sqrt{T/m}$. Therefore our diffusion result is qualitatively explained. Fuller analysis and more precise experiments give exact agreement.

Effusion

10. There is a simple case of diffusion, the treatment of which is not beyond us and which gives us an interesting application of the theory. It is the case of the flow of gas out through a small single hole into an evacuated space. Suppose once more that we have the box of Fig. 63, containing n molecules of gas at a pressure P, but now suppose a hole of area S has been punched in the side $BGEF$. How fast will the gas get out?

11. To simplify the treatment, we will suppose the box to be a cube with each edge 1 m. We have seen that the velocity component of the molecules parallel to OB is $W/3$ but that this is equally likely to be toward the side $BGEF$ or away from it. On the average then, half of the molecules will be moving toward $BGEF$ with a velocity $W/3$ at any instant. The probability that they will get out the hole in the side is the ratio of the area of the hole to the area of the side. But the area of $BGEF$ is one sq. m.; therefore the number getting out per second is

$$\frac{nWS}{6}. \tag{9}$$

The mass effusing per second is the number times the mass of each, but since the box has a volume of just 1 cu. m., $n\mu = \rho$, the density. Therefore the mass effusing per second is

$$\frac{n\mu WS}{6} = \frac{\rho WS}{6}. \tag{10}$$

But we saw that $W = \sqrt{3P/\rho}$; therefore the rate of effusion in kilograms per second becomes

$$\frac{S}{6}\sqrt{3P\rho} \tag{11}$$

if the pressure P is expressed in newtons per square meter and the density in kilograms per cubic meter. That the rate of effusion is proportional to the square root of the product of the pressure and the density is verified by experiment.

Distribution of Velocity

12. When we first introduced the ideas of the kinetic theory, we mentioned that the molecules did not all have the same velocity. We saw that this assumption was confirmed by the direct measurements of molecular velocities described in Chap. II, Pars. 24 to 27. Although we have shown in the present chapter that it is possible to deduce the most important of the gas laws in terms of the average value of the square of the velocity or its square root, we are naturally interested in more detailed information about the velocities of the molecules. Such information was first obtained by an application of the theory of probability to the problem. The theoretical analysis involved is

akin to that used in the study of vital statistics by actuaries. It is too involved for us to present, but we can at least give the results. These are as follows.

13. Suppose we have n molecules in a gas at an absolute temperature T. Let μ be the mass of a single molecule and k be Boltzmann's constant as before. Let c be the magnitude of the

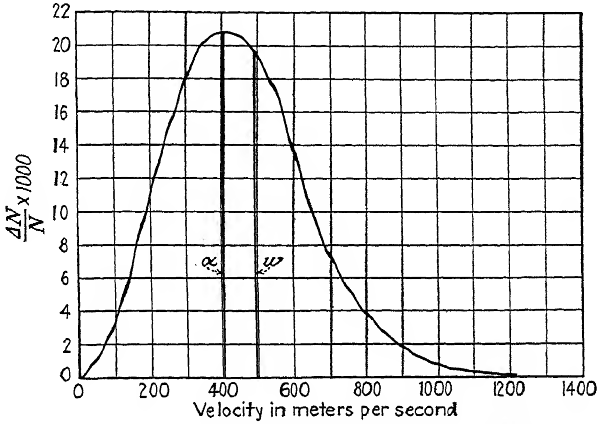


FIG. 65.—Distribution of velocity of the molecules of nitrogen at a temperature of 0°C. $\frac{\Delta N}{N} \times 1{,}000$ represents the number of molecules per thousand having a velocity within 10 m./sec. of the corresponding abscissa. α is the most probable velocity and W is the root mean square velocity.

velocity of a molecule. Then the number of molecules Δn, which have velocities between c and $c + \Delta c$, is given by

$$\Delta n = n \frac{4c^2 e^{-\frac{c^2}{\alpha^2}}}{\alpha^3 \sqrt{\pi}} \Delta c \qquad (12)$$

where $e = 2.718$ and is the base of the Napierian or natural system of logarithms and $\alpha = \sqrt{2kT/\mu}$. The significance of this equation is made more clear by reference to Fig. 65, which has been plotted on the assumption that $T = 273$,

$$\mu = 4.64 \times 10^{-26} \text{ kg.},$$

the mass of a nitrogen molecule, $n = 1{,}000$, and $\Delta c = 10$ m./sec. Evidently the number of molecules with small velocities is very

small. For example, the number having velocities between 10 m./sec. and $10 + \Delta c = 20$ m./sec. is too small to read on the curve. Calculation shows it to be 0.035 per thousand. If we go to higher velocities, say 200 m./sec., the curve shows that approximately eleven molecules per thousand will have velocities between 200 and 210 m./sec. The probability of a molecule having a velocity within 10 m./sec. of any given velocity increases to a maximum at a velocity just equal to α, in this particular case 403 m./sec. For this reason, α is called the most probable velocity. The probability of a molecule having a given velocity then begins to decrease but never vanishes. For the particular case considered, it is 68×10^{-6} at 1,200 m./sec. That is, if we had a million molecules instead of the thousand we chose, we might expect 68 of them to have velocities between 1,200 and 1,210 m./sec., about three times the most probable velocity.

14. We said that $\alpha = \sqrt{2kT/\mu}$ and now find that it is the most probable velocity; but the only velocity we have been dealing with heretofore was W, the root mean square velocity, which we saw equals $\sqrt{3kT/\mu}$. We see therefore that this is $\sqrt{\frac{3}{2}} = 1.224$ times as great as the most probable velocity and that this relation will hold for all gases and temperatures. In fact, the general shape of the curve is independent of these factors, although for higher temperatures or lighter molecules, α and W have larger values and the maximum of the curve will be farther to the right.

15. For the sake of completeness, perhaps we should mention the average velocity though it is not so useful as either the root mean square or most probable velocity. It can be shown to be equal to $2\alpha/\sqrt{\pi} = 1.128\alpha$ and therefore lies between the most probable and the root mean square velocities.

The Molar Heats of Gases

16. We have seen that each molecule of a gas has on an average a kinetic energy of $\frac{3}{2}kT$. Consequently if the temperature of a gas is raised, the kinetic energy of the molecules is raised a definite amount. For each degree rise in temperature, the average kinetic energy of a molecule must increase by $\frac{3}{2}k$ joules. If we are dealing with a mole of gas, the total energy that must go into increasing the kinetic energy of the molecules is $\frac{3}{2}Nk$ or $\frac{3}{2}R$ joules per degree rise in temperature where, as usual, N is

the number of molecules in a mole and R is the gas constant. We saw that the specific heat of a substance was equal to the amount of heat measured in kilocalories necessary to raise the temperature of one kilogram of the substance one degree centigrade. We shall now *define the molar heat of a substance as the heat per mole absorbed by the substance per degree rise in temperature.* It is evident that the molar heat is the same for all gases and is equal to $\frac{3}{2}R$ as long as the only change of the energy of the gas on heating is the change in the kinetic energy of the molecules. This is actually the case if the gas is monatomic and is kept at a constant volume during the heating. The molar heat of all monatomic gases at constant volume is, therefore,

$$c_V = \tfrac{3}{2}R = 12.45 \text{ joules/mole°C.}; \tag{13}$$

or, the energy being expressed in calories as is appropriate for a quantity of heat,

$$c_V = \tfrac{3}{2} \times 1.986 \text{ cal./mole°C.} \tag{14}$$

By definition, the specific heat, which we shall designate by C_V, is equal to c_V/M where M is the molecular weight of the gas. It is consequently not the same for different gases.

17. We have been careful to specify that we were talking about the molar heat at constant volume. We can also measure experimentally the molar heat at constant pressure. Let us see whether we can predict from our theory what it should be. Suppose a mole of gas at temperature T, pressure P, and volume V is in a cylinder closed by a movable piston P, of area A. Then $PV = RT$. Let the gas be warmed so that the temperature rises 1°C. Then if the pressure is kept constant, the gas must be allowed to expand to a new volume V' given by $PV' = R(T + 1)$. Therefore the change in volume is $V' - V = R/P$. But in increasing the volume, the piston is pushed out a distance d against the pressure P and this uses up energy. The amount of energy used is the product of the force PA on the piston and the distance d through which it moves, *i.e.*, PdA. But dA is obviously the increase in volume $V' - V = R/P$. Therefore the energy needed to expand the gas is

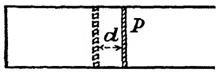


FIG. 66.—A cylinder with a movable piston P.

$$PdA = P(V' - V) = P\frac{R}{P} = R \tag{15}$$

and this energy must be supplied by the source of heat that is raising the temperature of the gas. Therefore the molar heat at constant pressure consists of two parts, $\frac{3}{2}R$ to speed up the molecules and R to move them apart, *i.e.*,

$$c_P = \tfrac{3}{2}R + R = \tfrac{5}{2}R. \tag{16}$$

Therefore the difference between the molar heats of a gas at constant pressure and at constant volume should be just equal to R, and their ratio should be $\frac{5}{3} = 1.67$. If the molecules are not rigid spheres as we have assumed, there may also be energy of internal motion of the parts of the molecules among themselves so that the molar heats may be greater than the theoretical values just calculated. The theory can be extended to cover this effect in diatomic gases and shows that then

$$c_V = \tfrac{5}{2}R \qquad \text{and} \qquad c_P = \tfrac{7}{2}R. \tag{17}$$

The excellence of the agreement with theory is shown in the following table.

TABLE 8.—MOLAR HEATS OF GASES IN R UNITS

	c_V	c_P	$c_P - c_V$	c_P/c_V
Monatomic gases:				
Theoretical	1.5	2.5	1.0	1.67
Helium	1.51	2.50	0.99	1.66
Neon	1.50	2.50	1.01	1.66
Argon	1.50	2.46	0.94	1.64
Diatomic gases:				
Theoretical	2.5	3.5	1.0	1.40
Air	2.56	3.60	1.04	1.41
Hydrogen	2.44	3.47	1.03	1.42
Polyatomic gases:				
Theoretical				
Carbon dioxide	3.63	4.84	1.21	1.33
Water Vapor	3.08	4.21	1.13	1.37

18. We have seen in this chapter that it is possible to develop a rigorous quantitative application of the kinetic theory to the behavior of gases. This theory explains the laws of Charles

and of Boyle, as it is intended to, and predicts the results of effusion and the distribution of molecular velocities found experimentally. Perhaps its most striking success is the prediction of the molar heats of gases described in the last paragraph. Further developments of the theory take account of the departures from the gas law shown by real gases, explain viscosity, the conduction of heat, the conduction of sound, and other properties of gases. The development of the theory for liquids and solids is much more difficult and has not yet been satisfactorily completed in quantitative fashion. Qualitatively the kinetic theory point of view is satisfactory in this field also. Some of the characteristics of substances in the liquid and solid states and the transitions between different states will be discussed in the next chapter.

SUMMARY

By assuming that the molecules in a gas are hard, perfectly elastic spheres in rapid random motion, the following equation is derived for the pressure exerted on the walls of a container:

$$P = \frac{n\mu\overline{W^2}}{3V} \quad \frac{\overline{W^2}\,\rho}{3}$$

where n is the number of molecules, μ the mass of one molecule, V the volume of the container, $\rho = n\mu/V$ the density and $\overline{W^2}$ is the mean value of the square of the velocities of the molecules. This equation is found to be equivalent to the general gas law [Eq. (11) of the last chapter] if $\frac{1}{2}\mu\overline{W^2}$, the mean kinetic energy of a molecule, is equal to $\frac{3}{2}kT$ where $k = R/N$ and is called Boltzmann's constant. Thus kinetic theory is reconcilable with experimental observations and makes a definite prediction about the mean kinetic energies of the molecules. Molecular velocities calculated from this prediction are consistent with those measured directly (Chap. II, Par. 25). The distribution of velocities among the molecules is also given by the theory.

One of the most striking successes of the theory is the prediction that the molar heats, the heats necessary to raise the temperature of one mole one degree, are the same for all monatomic gases. The molar heat at constant volume is $\frac{3}{2}R$ and at constant pressure $\frac{5}{2}R$. These predictions are verified by experiment.

ILLUSTRATIVE PROBLEMS

1. In a box there is a hole 1 cm.² in area. If the box is filled with oxygen at room temperature and 2 atm. pressure, how much mass will escape per second?

The rate of effusion in kilograms per second is $S\sqrt{3P\rho}/6$ where S is the area of the opening in the box, P is the pressure, and ρ the density of the gas in the box. Here $S = 1 \times 10^{-4}$ m.²,

$$P = 2 \times 0.76 \text{ m.} \times 13{,}600 \text{ kg./m.}^3 \times 9.8 \text{ m./sec.}^2 = 2.03 \times 10^5 \text{ newtons/m.}^2,$$

and the density is obtained from the perfect gas law

$$PV = m\frac{R}{M}T \qquad \text{or} \qquad P = \frac{m}{V}\frac{R}{M}T = \rho\frac{R}{M}T$$

whence

$$\rho = \frac{PM}{RT} = \frac{2.03 \times 10^5 \text{ newton/m.}^2 \times 32 \times 10^{-3} \text{ kg./mole}}{8.3\dfrac{\text{joules}}{\text{mole°C.}} \times 293\text{°C.}}$$
$$= 2.67 \text{ kg./m.}^3$$

Substituting these values in the equation for the rate of effusion gives

$$\frac{1 \times 10^{-4} \text{ m.}^2}{6}\sqrt{3 \times 2.03 \times 10^5 \frac{\text{kg. m.}}{\text{sec.}^2 \text{ m.}^2} \times 2.67 \text{ kg./m.}^3} = 0.0213 \text{ kg./sec.}$$

2. If 1 l. of hydrogen at 0°C. is compressed to one-half its volume at a constant pressure of 10 atm., how much heat must be removed from the gas? What is the final temperature? What would be the final temperature if hydrogen were a perfect gas?

The difference between the molar heats at constant pressure and constant volume multiplied by the change in temperature gives the work done by one mole of the gas in contracting. To obtain the total work, we multiply by the number of moles n.

$$n(c_P - c_V)Rt = P(V' - V)$$

where $c_P - c_V$ is in units of R and the work at constant pressure is the pressure times the change in volume. From this equation, we can find the final temperature. First we must find the number of moles of gas present from the perfect gas law $PV = nRT$. From this equation, converting to m.k.s. units (for pressure conversion see preceding problem or Table 26 of appendix) we have

$$10 \text{ atm.} \times 1 \text{ l.} = 10.13 \times 10^5 \text{ newtons/m.}^2 \times 10^{-3} \text{ m.}^3 = n \times 8.3 \text{ joules/mole°C.} \times 273\text{°C.}$$

or

$$n = \frac{10.13 \times 10^2 \text{ newton-m.}}{8.3 \text{ joules/mole} \times 273} = 0.447 \text{ mole}$$

The final temperature is then given by

$$0.447 \text{ mole} \times 1.03 \times 8.3 \text{ joules/mole°C.} \times t = 10.13 \times 10^5 \text{ newtons/m.}^2 \times (0.5 \times 10^{-3}\text{ m.}^3 - 10^{-3}\text{ m.}^3)$$

or

$$t = \frac{-10.13 \times 10^2 \times 0.5 \text{ newton-m.}}{0.447 \times 1.03 \times 8.3 \text{ joules/°C.}} = -133\text{°C.}$$

The heat that must be removed from the gas at constant pressure is the molar heat at constant pressure times the number of moles times the change in temperature.

$$H = nc_PRt = 0.447 \text{ mole} \times 3.47 \times 1.986 \text{ cal./mole°C.} \times 133\text{°C.} = 410 \text{ cal.}$$

If hydrogen were a perfect gas, the change in absolute temperature at constant pressure would be proportional to the change in volume from the perfect gas law so that since the volume is one-half, the absolute temperature is $^{273}\!/_{2} = 136.5$°K., or -136.5°C.

PROBLEMS

1. If 10^{17} molecules per second of hydrogen traveling 500 m./sec. strike 2 sq. cm. of a wall at an angle of 45°, how much pressure do they exert on the wall?

2. Compute the average speed of hydrogen molecules at NTP.

3. What is the "root mean square velocity" of the molecules of hydrogen at a temperature of 200°C.? At -67°C.? What is the most probable velocity and the average velocity of the molecules at the latter temperature? How large a proportion of the molecules have velocities between 200 and 210 m./sec.?

4. Ten grams of helium are in a 10-l. flask at a temperature of 20°C. The gas is escaping through a small hole 0.1 sq. mm. in area. How much will leak out in the first tenth of a second after the leak opens?

5. Carbon dioxide gas has the chemical formula CO_2. Find the fraction of molecules of CO_2 which at 70°C. have total speeds between (*a*) 100 and 200 m./sec. and (*b*) 400 and 600 m./sec.

6. If the flask in Prob. 4 has no leak and is raised to a temperature of 100°C., what is the pressure? If the flask alone has a heat capacity of 150 cal./deg., how much heat was required to raise the temperature of the gas and the flask to the higher temperature?

7. N molecules of nitrogen at 0°C. have the usual average velocity perpendicular to one face of a box of $\frac{1}{2}$ sq. m. area. There is a hole 2 sq. cm. in this face. What is the mass effusing per second at room temperature?

8. How many molecules of air (assume molecular weight = 30) escape from an automobile tire per second through a valve 3 sq. mm. in cross section at a pressure of 30 lb./sq. in. gauge and 50°C.?

9. What is the kilogram molecular heat of a monatomic gas?

10. Two compound gases A and B have molecular weights 99 and 81, respectively. Gas A has a specific heat at constant pressure of 0.23 and

gas B has a specific heat at constant pressure of 0.084. Calculate the molar heats at constant pressure, and state whether these gases are diatomic or polyatomic.

11. An air-conditioning apparatus draws in 1,000 l./sec. from the atmosphere outside. If the outside temperature is 95°F. and the inside temperature is 70°F., how much heat does the machine remove per second?

12. What is the temperature in the center of a completely evacuated box whose walls are red hot (600°C.)?

13. How much work can be got out of 10 g. of water vapor by adding 100 cal. of heat at constant pressure? What is the rise in temperature?

14. If H cal. of heat are added to a perfect gas at constant pressure P newtons/sq. m., what will its change in volume be?

15. How much heat must be added to 50 g. of carbon dioxide at constant pressure to do 100 joules of work in expansion?

16. How much heat must be removed from air at constant pressure in order to have 10^4 joules of work done by it?

17. If 100 joules of work are done by 50 g. of argon in expanding at constant pressure, what is the rise in temperature?

18. One hundred grams of hydrogen are heated by 1,000 cal. at constant pressure. What is the difference between the final temperature of the hydrogen and the final temperature if hydrogen were a perfect gas?

19. Ten grams of helium are heated by 100 cal. at constant pressure. If the gas is originally at NTP, what is the final volume?

CHAPTER IX

CHANGE OF STATE

1. In the last chapter we saw that it was possible to apply the laws of mechanics to the assumptions of the kinetic theory of matter. But in making this application we limited ourselves to the case of gases, in fact to the case of an ideal gas, where the forces of attraction between the molecules and the finite size of the molecules could be neglected. This limitation enabled us to carry through a complete calculation which we then were able to apply to the behavior of gases at low pressure. The success of this treatment gives us confidence in the correctness of our general view that all matter is made up of tiny particles in chaotic motion and that the amount of this motion increases with rising temperature. In the present chapter we want to consider the interplay between the disruptive effect of this thermal motion and the cohesive effect of the forces of attraction between the molecules. We want to consider not only gases but also liquids and solids and in particular the transitions between these different states or phases of matter as the temperature is changed. Although we shall find the problem too complex to handle with any degree of mathematical precision, the ideas developed in the last chapter will continually be called upon.

2. Let us review our general ideas about the nature of liquids and solids. We assumed that there were forces of attraction between molecules which varied rapidly with the distance so that the attraction between molecules that were more than a few ten-millionths of a millimeter apart was very small but became considerable when the molecules approached more closely. We found that the motion of the molecules varied with temperature and looked as if it might disappear entirely at the absolute zero. At the absolute zero then, once molecules are brought together, there is nothing to oppose the attractive force and they will stay close packed. In other words, all substances must be solids at the absolute zero. This expectation has been justified experi-

mentally by showing that even helium freezes at a temperature of 1.13°K. and a pressure of 25.3 atm. As the temperature increases, the molecules begin to move. All the molecules do not have exactly the same amount of motion. Their kinetic energies are distributed according to the same law that holds for the kinetic energy of gas molecules as illustrated in Fig. 65, page 175. Thus although most of the molecules are merely vibrating about positions of equilibrium a few of those in the surface layer of the substance will occasionally break away into free space and behave as if they were in the gas phase. The main body of the substance is a solid, but there is some evaporation from the surface. Then, as the temperature rises, the vibration of the molecules increases. The evaporation from the surface increases, and finally the average energy of vibration becomes so great that the whole rigid structure of the solid breaks down and the molecules start moving around in a much more chaotic fashion, though still remaining close to each other and within the distance where the attractive force is great. This is the liquid state. Usually the volume increases as a result of liquefaction. With ice and a few other solids, however, the volume decreases, oddly enough. (It is interesting to speculate as to what our climate would be if this were not the case for ice.) Finally the tendency to evaporate becomes so great that all the molecules get away from each other forming a gas. The point at which this happens will obviously depend on the ease with which molecules that have evaporated can stay away from the others. If they are in a confined space so that they tend to bounce back into the liquid, the formation of the gas will be retarded. In other words, this process will depend on the pressure, as will the melting process also.

3. The change from solid to liquid to gas is the most familiar effect of rising temperature, but we have seen that there are always some molecules escaping as vapor. In some cases this tendency increases so rapidly that the substance goes directly from solid to gas without passing through the liquid state at all. This process is called sublimation. Examples are furnished by solid carbon dioxide (dry ice) and by iodine which at atmospheric pressure do not liquefy at all as the temperature is raised but change directly into vapor.

Vapor Pressure

4. If a straight glass tube is filled with mercury and then inverted with its open end under the surface of a reservoir of mercury, it forms an ordinary barometer in which the enclosed mercury will stand at a height of 760 mm. above that of the surface of the reservoir. This is, of course, merely a measure of atmospheric pressure. Now if a drop of water is inserted under the submerged end of the tube it rises to the space at the top of the tube. The upper surface of the mercury column is seen to be pushed down to a lower position than before. If two other similar tubes are set up beside the first one and drops of alcohol and ether introduced into them, the mercury is seen to be standing at still lower levels as shown in Fig. 67. Now if one of the tubes is pushed down into the mercury, the amount of liquid is seen to increase somewhat but the height of the surface remains approximately constant. Similarly if the tube is raised, the amount of liquid decreases but the height remains constant.

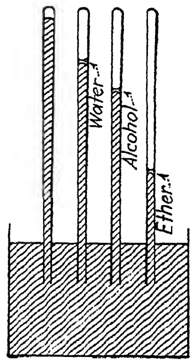


FIG. 67.—Pressures of saturated vapors. The vapor pressure is roughly given by the difference in height between the barometer on the left and the other mercury columns.

5. We can understand what is happening in terms of kinetic theory quite easily. Before the introduction of the liquid, the space above the mercury column is empty (save for a few mercury molecules). When the liquid, which is lighter than the mercury, is introduced, it rises to the top and some of its molecules escape from the surface into the empty space. This continues until the space has so many molecules knocking around in it that just as many are returning to the liquid as are evaporating from it. The liquid and its vapor are then said to be in equilibrium, and the vapor is said to be saturated. Since the molecules of the vapor are continually bombarding the liquid surface, just as the molecules of a gas bombard the surface of its containing vessel, they exert a pressure and push down the liquid and the mercury under it. The tendency of the alcohol molecules to evaporate is greater than that of the water and that of the ether is greater still so that the mercury in the three tubes stands as shown in the

figure. Now if a tube is pushed down, the volume of the space available for the vapor is diminished so that fewer molecules can get into it before there are as many getting bounced back into the liquid as are getting off. Consequently the proportion of liquid to vapor increases. If, on the other hand, the tube is raised so that the volume occupied by the vapor increases, more and more liquid has to evaporate to fill the space to equilibrium pressure. If the volume increases sufficiently, all the liquid evaporates and the pressure begins to fall. The vapor then begins to behave like a gas obeying Boyle's law at least to a first approximation.

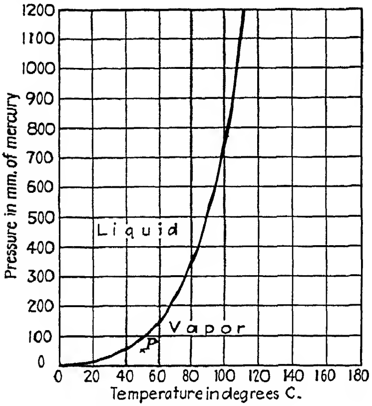


FIG. 68.—Vapor pressure of water at intermediate temperatures.

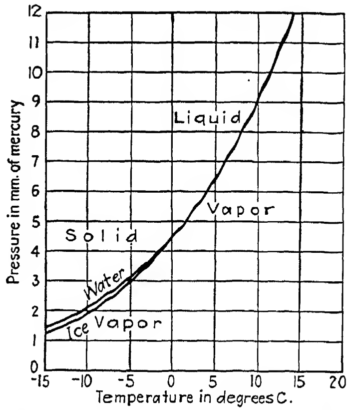


FIG. 69.—Vapor pressure of water and ice at low temperatures.

6. The difference in the height of the mercury column before and after the introduction of the liquid, when corrected for the weight of the liquid itself, is a measure of the vapor pressure of the liquid. It is no surprise to find that this increases with rising temperature. The curve in Fig. 68 shows how the vapor pressure of water varies with the temperature between 0°C. and 120°C. This is a perfectly smooth curve that can be fitted by an empirical formula and it is possible to extend it below the freezing point, *i.e.*, to calculate the vapor pressure of water cooled below the freezing point but not solidified. These calculations can be verified by measuring the vapor of supercooled water. The vapor pressure of ice in this same range of tempera-

ture is not quite the same as can be seen from the curves shown in Fig. 69. This is not surprising since the intermolecular forces in ice that oppose evaporation might naturally be expected to differ from those in water. It is perhaps more surprising that the vapor pressures are so nearly the same.

Dalton's Law of Partial Pressures

7. In 1802, Dalton found as an experimental result that

The pressure of a mixture of gases and vapors that do not react with each other chemically is the sum of the pressures which each would exert if it alone were occupying the same volume occupied by the mixture.

Like most of the gas laws we have been discussing, this law is no longer exact if the pressure is very great. In terms of kinetic theory, the law holds as long as the space not occupied by molecules is large compared with the space occupied by the molecules themselves. The law might, in fact, be deduced from the kinetic theory since there is no more reason why the action of a molecule should be influenced by the presence of another molecule of a different type than by the presence of another molecule of the same type. Our derivation of pressure in Chap. VIII assumes no mutual influence between the molecules.

8. Applying Dalton's law to the phenomena of vapor pressure that we have been describing, we conclude that the mercury column in the barometer tube would have been pushed down just as much by the introduction of a drop of water if there had been air above the mercury to start with. This is correct. The vapor pressure of a liquid is independent of the presence of a gas over the liquid. It should be pointed out, however, that the presence of a gas over a liquid will greatly retard the process of evaporation even though the final equilibrium vapor pressure will be the same. This fact was illustrated by the iodine-vapor experiment of Chap. II.

Heat of Vaporization

9. Since the molecules in the solid or liquid are held together by forces which have to be overcome before vapor can be formed, it is clear that it requires energy to convert solid or liquid into vapor. Work must be done to move two molecules apart, just as work must be done to move two ends of an elastic band apart.

The only source of energy available is the heat energy of the molecules themselves. Consequently any evaporating liquid cools itself and its surroundings unless some source of external heat is supplied. This is a phenomenon that is continually encountered in nature. In fact, the even temperature of the human body is maintained by control of the amount of perspiration, the evaporation of which cools the body. If cooling by evaporation did not occur, it would be impossible for human beings to endure external temperatures higher than the normal temperatures of the body. It is of course particularly effective in dry climates. For this reason, high temperatures in dry regions can be borne with far less discomfort than in humid regions. In such arid regions water for drinking purposes is often cooled merely by storing it in porous containers of skin or crockery, so that part of it seeps through and evaporates.

10. The amount of energy per unit mass required at constant temperature to convert a liquid into a vapor is called the *heat of vaporization* of the liquid. It depends, of course, on the nature of the liquid and it also depends somewhat on the temperature at which the vaporization takes place. Some values of this quantity are given in Table 9.

TABLE 9.—HEATS OF VAPORIZATION

Substance	Temperature °C.	Heat of vaporization, kg.-cal./kg.
Water	0.0	595.9
	50.0	568.5
	100.0	539.6
	150.0	503.5
Carbon dioxide (solid)	− 60.0	87.2
Ether	34.6	83.9
Alcohol	78.4	204.0

For additional values, see Table 9 in the Appendix.

Equilibrium and Equilibrium Diagrams

11. In studying mechanics, we speak of a system as in equilibrium under the action of various forces if the positions of the various parts of the system do not change with time. Similarly in dealing with changes in the phase or state of substances, we

say that a system is in equilibrium if the proportions of the different phases and the temperature and pressure of the substances do not change with time. We are confining ourselves to systems of one component, *i.e.*, systems where all the phases have the same composition. We shall not attempt, for example, to deal with the question of the vapor pressure or the freezing point of a solution. This allows us to avoid making a general definition of what we mean by "phase" since in the cases we are considering it will be simply equivalent to "state," *i.e.*, whether the substance is solid, liquid, or gaseous. Leaving the more general problems of this field to the books on physical chemistry, we shall consider the particular problem of equilibrium between the various states or phases of a pure substance.

12. Returning to the vapor-pressure curve of Fig. 68, we see that points on it represent conditions of equilibrium between the liquid and vapor states for water (if we may be allowed to use the word water for the chemical compound H_2O in whatever state it is). The two variables that we are considering are pressure and temperature. Heretofore we thought of this curve as giving us the pressure of the saturated vapor at any given temperature. Now we want to think of it from a slightly different point of view as giving us the only combinations of temperature and pressure that represent conditions of equilibrium between the liquid and the vapor. Suppose, for example, we take some point off the curve such as the point at 50° and 50 mm. marked P in Fig. 68 and consider what happens to a sample of water put in a vessel maintained at that temperature and pressure. Since the vapor pressure of the water at 50° is over 90 mm., more molecules will leave the liquid than will return to it. The pressure will tend to build up, but we are assuming it is held constant (by increasing the volume) so that the liqu d will continue to evaporate until none is left. We shall then have water vapor acting like a gas at a temperature of 50° and a pressure of 50 mm. In other words, we cannot have both liquid and vapor existing in equilibrium at this temperature and pressure, we can have only the vapor at this temperature and pressure. This is true for all points on the pressure-temperature diagram below the vapor-pressure curve. For points above the curve, on the other hand, a similar argument shows that only the liquid state can exist since the pressure is always greater than the vapor pressure so that the vapor is all

condensed. Looking at it a little differently, if we start with some water and water vapor at a p and t off the curve and let the combination alone, evaporation or condensation will take place until there is nothing but liquid, nothing but vapor, or the pressure and temperature have changed enough to come on to the curve. Such processes have already been described in terms of molecular motion in Par. 5.

13. The vapor-pressure curve of Fig. 68 runs down only to zero, but we have seen that ice also has a vapor pressure, the curve of which is the lower of the two curves in Fig. 69. This curve then gives the equilibrium condition between ice and water vapor in this region of temperature. If we try to keep some ice at a temperature and pressure corresponding to a point below the curve, it will all sublime into vapor. If we leave it alone, enough will evaporate to bring the pressure up and the temperature down until the curve is reached. We have here an equilibrium curve between the solid and vapor phases corresponding to the previous equilibrium curve between the liquid and vapor phases.

14. We have seen that if water vapor at a fairly high temperature, say 50°C., is compressed, water is formed, but if the vapor is at a low temperature, say -15°C., and is compressed, ice is formed. In short, we have studied the equilibrium between the solid phase and the vapor phase and between the liquid phase and the vapor phase. We still have to consider the transitions from the liquid to the solid phase, and vice versa, *i.e.*, freezing and melting. Suppose we start with some water at a temperature of 50° and a pressure of 200 mm. From Fig. 68, we see that it must all be liquid. Now lower the pressure to 1.0 mm., keeping the temperature constant. The water will all evaporate, changing to vapor. Now keep the pressure constant but lower the temperature to -10°. We are still below the vapor-pressure curve (Fig. 69) so that there is still nothing but vapor. Now keep the temperature constant, but raise the pressure to 200 mm. again. This brings us to a point far above the ice-vapor equilibrium curve and therefore converts all the water vapor to ice. Now finally to complete the cycle, let the temperature rise again to 50°, bringing us back to our starting point. Somewhere in the course of this last temperature rise, ice melts to water, and there must be some point where the two phases exist in equilibrium for this particular pressure of 200 mm. The temperature at this

point is the melting point of ice for this particular pressure. Now it so happens that the variation of the melting point with pressure is comparatively small so that the ice-water equilibrium curve is almost vertical at low pressures in contrast to the ice-vapor and vapor-water equilibrium curves. The curve is given in Fig. 70 and like the other curves can be thought of in two ways, either as simply the variation of the melting point with pressure or as the ice-water equilibrium curve. It can be seen that the variation of the freezing point with pressure amounts to only about 0.01°C./atm.

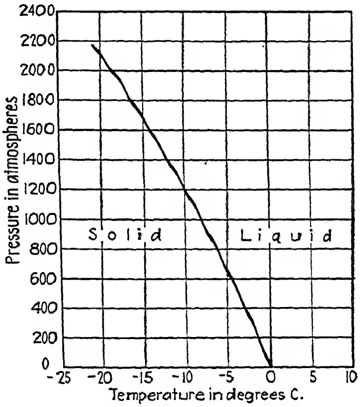


FIG. 70.—Variation of the melting point of ice with pressure. The graph shows the ice-water equilibrium curve. (Note the magnitudes of the pressures.)

Regelation

15. That the freezing point really is lowered by pressure is shown very nicely by an experiment that is often performed in lecture. A block of ice is supported at each end and a piece of wire looped around its middle. A weight of several kilograms is then hung from the wire so that it is under tension. The wire gradually sinks into the ice and will eventually pass entirely through it but the ice above the wire freezes solid again (regelation) so that the wire passes through the block without breaking it. The interpretation of this experiment is left to the student. Another example of the same thing is the packing of a snowball into a single solid mass or the formation of ice from snow in a glacier.

Heat of Fusion

16. We saw that it took a certain amount of energy to convert a liquid to a vapor. The change of the arrangement of the molecules of a substance when it melts is not so radical but it is nevertheless considerable and also absorbs energy. This energy is called the heat of fusion of a substance. Like the heat of vaporization, it depends somewhat on the temperature where the

fusion occurs but this dependence is not so great as in the case of the heat of vaporization. For practical purposes, it can be considered constant. The heat of fusion of water at zero is approximately 80 cal./g. This is also sometimes called the latent heat of fusion of water.

TABLE 10.—HEAT OF FUSION

Substance	Kg.-cal./kg.	At °C.
Carbon dioxide	45.30	− 56.2
Copper	43.2	1083
Ethyl alcohol	24.9	− 112
Iron	48.0	1535
Lead	5.86	327
Tin	14	231.8
Water	79.67	0
Zinc	26.1	419.4

17. If it takes heat to melt ice, one might expect the freezing of water to give off an equal amount of heat, and indeed it does. Similarly, the condensation of a vapor gives off heat. These absorptions and emissions of heat that accompany the changes in state of substances are of great help in reducing the temperature fluctuations that occur in nature.

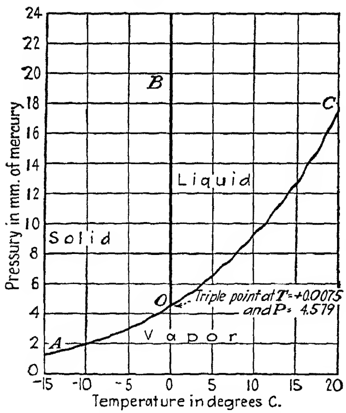


FIG. 71.—Equilibrium diagram for water (to scale).

Complete Equilibrium Diagram. The Triple Point

18. We have considered the pressure-temperature equilibrium conditions for three states of water taken in pairs but there is no reason why the whole situation cannot be represented in one diagram. In fact, we bring out one very interesting point by doing so although we cannot conveniently cover so great a range of pressures and temperatures as was included in the previous diagrams. In Fig. 71, the results of Figs. 68, 69, and 70 are com-

bined. The graphs are drawn quantitatively to scale as shown by the coordinates. This has the disadvantages that it is hardly apparent that *AO*, the vapor-pressure curve for ice, is not simply the continuation of *CO*, the vapor-pressure curve for water, and that on this scale of coordinates the decrease of the melting point with pressure does not show up at all, the curve *OB* appearing parallel to the pressure axis. We have therefore redrawn the diagram in a schematic way, not to scale, in Fig. 72.

19. The point *O* where the three curves meet is called the triple point and represents a condition of pressure and tempera-

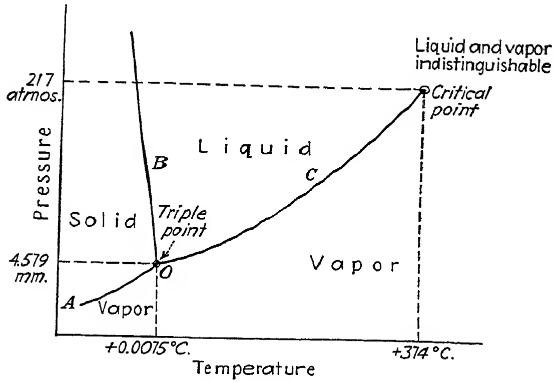


FIG. 72.—Equilibrium diagram for water (schematic).

ture where all three phases vapor, liquid, and solid can exist simultaneously in equilibrium. Perhaps its significance can be brought out most clearly by making an imaginary experiment. Suppose we could introduce a piece of ice at a temperature below zero into an evacuated vessel that is not perfectly heat insulated. Our system would then be starting from a point below *AO* and would not be in equilibrium. The ice would sublime until the vapor pressure had built up enough to bring the system into equilibrium somewhere on the line *AO*. As heat gradually leaked in, it would supply energy enough to raise the temperature of the ice and vaporize more of it so that the system would move slowly up the line *AO* until it reached *O*. At this point, the ice would begin to melt as well as to vaporize and the pressure and temperature would remain constant until enough

heat had leaked in to melt all the ice. During this period, there would be ice, water, and water vapor present simultaneously. As still more heat leaked in, the system would move up the curve *OC* with the ice all melted and the water and vapor in equilibrium. If the supply of heat were cut off when the system was at *O* the ice, water, and vapor would remain in equilibrium indefinitely. If in this condition a piston was gradually pushed down, decreasing the volume, the vapor would all condense. If the push of the piston on the water-ice mixture were still further increased, the ice would all melt since the system would be forced up into the region between *OB* and *OC*. If some heat were allowed to leak off, the system would tend to follow up the curve *OB* with ice and water in equilibrium. In Par. 29, we will describe an experiment that is very similar to this imaginary one and that can actually be tried in lecture.

20. Before going on to a discussion of the critical temperature plotted at the upper end of the *OC* curve, we should mention that the equilibrium diagram for water as we have shown it is by no means complete. Studies at high pressure, particularly by Professor Bridgman of Harvard, have shown that there are several different kinds of ice formed at lower temperatures and higher pressures. A complete pressure-temperature equilibrium diagram would show the transition lines and triple points for these also but such complications are rather beyond us at present. It might also be mentioned that the liquid-solid equilibrium line for water is different from that for most substances. In most substances, the solid form is more dense than the liquid, in contrast to water; and for such substances the freezing point increases with increasing pressure so that the liquid-solid equilibrium line makes an angle of less than 90° with the temperature axis.

Critical Temperature

21. The curve of Fig. 73 gives the vapor pressure of water at high temperatures. Suppose we consider the conditions that it shows hold in the neighborhood of 350°. If the pressure is such that we are at a point just below the curve, the water is all in the vapor state but it is highly compressed so that the molecules are very much closer to each other than at atmospheric pressure and zero temperature. We cannot calculate the density exactly because the vapor certainly will not obey the laws for a perfect

gas under these conditions. But if we apply these laws, we at least get some idea of the density. Assuming an absolute temperature of 620° and a pressure of 200 atm. and that the gas laws apply, we find the density of the water vapor to be 0.07 g./cc. or about one-tenth the density of water at room temperature. The actual density will be somewhat larger since the attractive forces between the molecules will certainly make themselves felt at this density and tend to bring the molecules closer together. On the other hand, the water phase which we still reach if we move

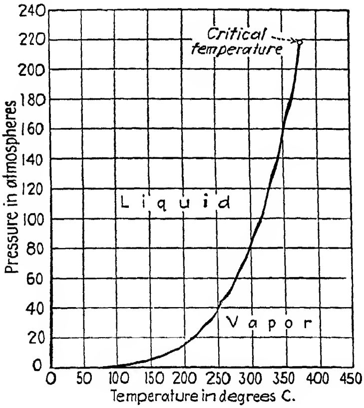


FIG. 73.—Vapor pressure of water at high temperatures.

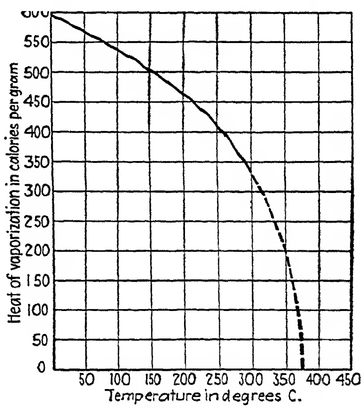


FIG. 74.—Variation of the heat of vaporization of water with temperature.

upwards to just above the curve will have a density nearly the same as that of water at ordinary conditions since the thermal expansion will counteract, probably more than counteract, the compression caused by the high pressure. The difference in density in the two phases is therefore only of the order of a factor of 10. Since density is proportional to the number of molecules per unit volume and this is proportional to the cube of the average distance between them (see Chap. II, Par. 21), this distance is increased by a factor of only two or three in changing from liquid to vapor state at this temperature. This is a very much less radical change than that which occurs in evaporation slightly above the zero. The density of the vapor just below the equilibrium curve at 5° is about 8×10^{-6} g./cc. so that for evaporation at this temperature the average distance between the molecules

increases by a factor of about 200. We might expect from this reasoning that the amount of energy needed to go from the liquid to the vapor phase might be lower at high temperatures than at low and this is in fact the case as can be seen from the graph of Fig. 74 which shows the heat of vaporization of water as a function of temperature.

22. The discussion of the last paragraph shows that as we follow the vapor-liquid equilibrium curve of water up to very high temperatures the liquid and vapor phase become more and more alike. The question naturally arises as to whether they ever become indistinguishable, and the answer is that they do. This can be shown very prettily by an experiment with carbon dioxide. A heavy sealed glass tube is filled with carbon dioxide at such a pressure that at room temperature the tube is about half filled with liquid. If it is warmed somewhat, the meniscus marking the division between the liquid and the vapor suddenly disappears. This is not caused by the gradual evaporation of the liquid since the meniscus is clearly visible near the middle of the tube just before it disappears. The effect is the one we have suggested, the vapor and the liquid have become equally dense so that their properties are identical and no division between them is possible. Strictly speaking, it is not proper to call the resulting assemblage of molecules either a liquid or a vapor; but since the thermal agitation of the molecules in it predominates over the molecular attraction and this effect becomes more striking as the temperature is raised still further, it is usually stated that above this point the substance is gaseous. This is also in agreement with the notion that a true liquid always shows a free surface.

23. The experiment just described as performed with carbon dioxide can be done with other substances and, indeed, could be performed with water if a glass tube that was strong enough to sustain the necessary pressure could be made. The point on the vapor-liquid equilibrium curve where the densities of the two phases become identical, and the heat of vaporization zero, is called the critical point and the corresponding temperature and pressure are called the critical temperature and critical pressure. The surface tension of the liquid also approaches zero as this point is approached. The significance of the critical point will be made more clear by a discussion of the pressure-volume

relations for a real substance which will be given in the next paragraphs.

Isotherms and the Critical Temperature

24. In the experiments we have been describing, we have tacitly assumed that the volume was kept constant and then have studied the relation between the pressure and the temperature. Now we want to consider how the pressure varies with the volume for a given amount of substance when the temperature is held constant. We know that for a perfect gas the relation $PV = RT$ holds and the pressure-volume curve is a hyperbola. Two such curves are shown in the upper part of Fig. 75. But we want to investigate the actual observed behavior of a substance like carbon dioxide which we know is not a perfect gas. Suppose we take a certain amount of the gas in a cylinder which is closed with a piston and whose temperature can be controlled. Then suppose we observe the volume occupied by the gas as the pressure on the piston is increased. If this experiment is performed at a temperature of 80 or 90°C., we get a curve that appears to be a smooth hyperbola like those obtained for a perfect gas. Since all the observations plotted on this curve are taken at the same temperature, the curve is called an isotherm. If we take other series of pressure-volume observations at other temperatures, we get a whole family of isotherms such as are shown in Fig. 75 and we observe that the shape of the isotherms changes as the temperature gets lower. Furthermore, if the gas in the cylinder is watched, it is found that the sudden changes in the trend of one of the low-temperature isotherms correspond to changes in the phases of the carbon dioxide in the cylinder. Take, for example, the isotherm for 21.5°C. If we start with large volume and low pressure, increasing the pressure on the piston decreases the volume gradually until the gas begins to liquefy. We then have, obviously, the condition of a saturated vapor in equilibrium above its liquid; and therefore the slightest increase in the pressure will decrease the volume until all the vapor has condensed. The isotherm, therefore, will run parallel to the volume axis until all the vapor is condensed and the space under the piston filled completely with liquid. Any further decrease in the volume can be achieved only by the elastic compression of the liquid. This requires large changes in pressure for very small

changes in volume so that the isotherm rises rapidly almost parallel to the pressure axis.

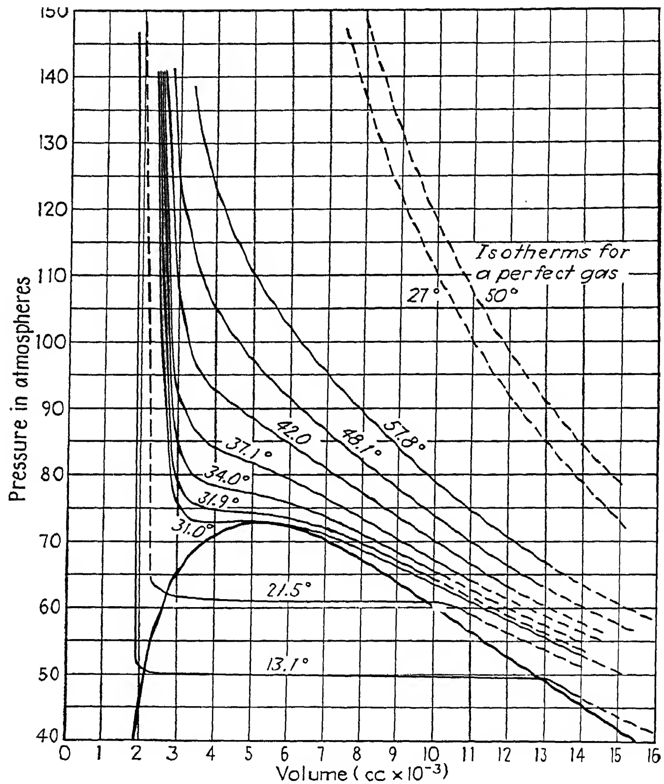


FIG. 75.—Isothermal pressure-volume curves for an amount of carbon dioxide that occupies one cc. at 0°C and 1 atm. Temperatures are in degrees centigrade. Liquid is formed only under the dome-shaped curve. The top of the dome is the critical point at 31°C. No liquid is formed above this temperature.

25. As the temperature is raised, the horizontal part of the corresponding isotherm becomes shorter until we finally get to an isotherm that shows a point of inflection but no horizontal part. At still higher temperatures, even the point of inflection vanishes as we have seen. Suppose we fix our attention on the isotherm in

which the horizontal part of the curve just vanishes. Evidently at this temperature no liquefaction occurs since there is no corresponding region of diminishing volume without change of pressure. This indirect conclusion is confirmed by the direct observation that no liquid appears in the cylinder at this temperature no matter how great the pressure is made. None of the isotherms above this one show horizontal portions, but for every lower isotherm there is a particular pressure at which the curve flattens out and liquid appears. It is evident, therefore, that the gas can not possibly be liquefied at a temperature higher than that where the horizontal part of the curve vanishes, and can be liquefied by sufficient pressure at any lower temperature. This temperature is therefore called the *critical temperature* and the corresponding pressure the *critical pressure.* The point on the diagram where the critical temperature isotherm just touches the region where vapor and liquid exist in equilibrium is called the *critical point.* The student will find it useful to compare this method of arriving at the meaning of the critical point with that given in connection with Fig. 73. The critical data for a number of substances are given in Table 11.

TABLE 11.—CRITICAL DATA

Substance	Critical temperature, °C.	Critical pressure, atm.	Density, g./cc.
Air	−140.7	37.2	0.31
Argon	−122	48	0.531
Benzene	288.5	47.7	0.304
Carbon dioxide	31.1	73.0	0.460
Carbon disulfide	273	76	
Carbon monoxide	−139	35	0.311
Chlorine	144.0	76.1	0.573
Ethyl alcohol	243.1	63.1	0.2755
Helium	−267.9	2.26	0.066
Hydrogen	−239.9	12.8	0.0310
Methane	− 82.5	45.8	0.162
Neon	−228.7	25.9	0.484
Nitric oxide	− 94	65	0.52
Nitrogen	−147.1	33.5	0.311
Oxygen	−118.8	49.7	0.430
Sulfur dioxide	157.2	77.7	0.52
Water	374.0	217.72	0.4

The exceedingly low values of the critical temperatures of such gases as nitrogen, oxygen, hydrogen, helium, etc., show why early attempts to liquefy them were unsuccessful and why they came to be called "permanent gases."

Freezing Point and Boiling Point

26. It may have seemed strange that in our elaborate discussion of the changes of state of water no specific mention was made of the familiar freezing point at 0 and boiling point at 100°C. This was because in a sense these points are accidental since they are determined by atmospheric pressure. The equilibrium diagram we discussed was for water in an evacuated vessel. If we have to consider water in contact with air at atmospheric pressure, the situation is more complicated; but for our purposes, we may consider that the effect of the air is merely to raise the pressure 1 atm. Remembering the small effect that pressure has on the freezing point, we see why it is that zero, the freezing point under atmospheric pressure, is only 0.0075° below the triple point.

27. Boiling is a process that we have not considered as yet. It is simply the phenomenon that occurs when the temperature in a liquid is so high that the vapor pressure is equal to the hydrostatic pressure in the liquid. Small pockets of vapor then form in the body of the liquid itself and being of smaller density than the liquid rise to the surface. The reason that they usually form on the bottom, in practice, is because this is where the heat is supplied. In spite of the fact that the pressure is somewhat greater than nearer the surface, the temperature rise is usually so much faster that the boiling point is reached sooner. At the surface of the water, the pressure is simply the pressure of the atmosphere and, therefore, boiling occurs just at the surface when the vapor pressure is just equal to 760 mm. Consultation of the vapor-pressure curve of Fig. 68 shows that this is at 100°C. as it must be since this was the way 100°C. was defined on our temperature scale. It is obvious from this explanation that the boiling point will vary with pressure. At a high altitude above the sea, the pressure of the atmosphere is less and, therefore, water will boil at a lower temperature. The boiling point for any particular pressure can be read directly

from the vapor-pressure curve. An interesting natural example of the raising of the boiling point by pressure is the geyser.

28. Although boiling in a sense is nothing but evaporation, it is so rapid compared with ordinary evaporation that a boiling liquid has different characteristics from one that is merely evaporating at the surface. One of the most useful of these characteristics is approximate constancy of temperature. If a kettle of water is put over a hot fire, the fire will supply more energy than is required for the vaporization that takes place at the surface. The water, therefore, will rise in temperature until it reaches the boiling point. As soon as boiling begins, vapor is being formed at many points throughout the water, the amount of vapor being formed per second is enormously increased, and all the heat that the fire can supply is taken up. In other words, boiling tends to cool water. If the supply of heat from the fire should go down, parts of the water would drop very slightly below the boiling point, the amount of vaporization would decrease, and the average temperature of the water in the vessel would still be very nearly 100°. This property is invaluable in the boiling of eggs and other thermochemical processes.

Simultaneous Boiling and Freezing

29. The dependence of the boiling point on the pressure and the cooling effect of boiling can be shown in a rather striking experiment. Suppose we put some cold water on a watch glass under a bell jar connected to a vacuum pump. It is also desirable to have a small beaker of sulfuric acid under the bell jar to absorb the water vapor. If the pump is started, the pressure is reduced and the water is cooled by evaporation. The system is of course not in equilibrium, but if the pumping is fast the pressure will fall so rapidly that the water will boil. As long as the boiling is occurring, the pressure and temperature must correspond to a point on the curve OC of Fig. 72. The effect of the boiling is to keep the temperature and pressure on this curve. If the vapor is momentarily pumped away too fast, the pressure falls, and the system starts to go into the region below OC; but then the boiling becomes more vigorous, producing more vapor and cooling the water, thereby tending to bring the system back onto the curve farther to the left. Sometimes if the evacuation is

not rapid enough, the water ceases to boil but is still cooled rapidly. In this case, the system cuts over from the curve OC to the curve OB and suddenly all the water freezes. However, with the proper adjustment, it is possible to bring the system down along the curve OC toward the triple point so closely that the water can be seen to solidify while it is still boiling.

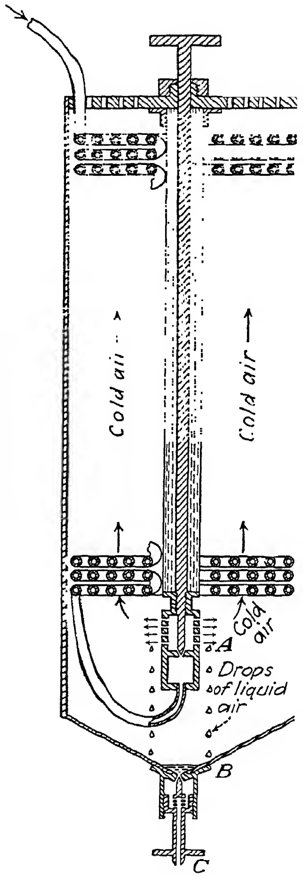


FIG. 76.—Precooling coil and expansion valve for a liquid-air machine. The compressed air enters the coil at the top and expands through the valve A at the bottom.

Liquefaction of Gases

30. As we have already mentioned, modern low-temperature technique has succeeded not only in liquefying but in solidifying all gases so that the old-fashioned term "permanent gases" has lost all meaning. As can be seen from Table 11, very low temperatures are necessary for the liquefaction of such gases as air, hydrogen, and helium, the old "permanent gases." Conversely, the boiling points of these gases under atmospheric pressure are very low, and therefore very low temperatures can be maintained by the use of them in the liquid state. The use of liquid air for various purposes is now very extensive. It is of interest therefore to consider how it is produced.

31. The liquid-air plant in Palmer Laboratory at Princeton may be taken as typical of one of the two common methods of producing liquid air. The principle it uses is the cooling of air by its own expansion. As we have already pointed out, there is an attractive force between the molecules that is appreciable even in the gaseous state so that work has to be done to increase the

average distance between them. If the air is allowed to expand suddenly, this work is done at the expense of the thermal energy of the molecules and therefore the temperature drops. In practice, the amount of expansion needed to get a sufficient fall in temperature is very great so that the chief part of a liquid-air plant is the compressors. In the one at Princeton, the air is compressed in four stages to a pressure of about 300 atm. The heat developed in this compression is carried off by water cooling. The air at this pressure and at a temperature slightly below that

FIG. 77.—A photograph of the liquid-air plant in Palmer Laboratory.

of the room is then led into the liquefier which is shown in the diagram in Fig. 76. This consists essentially of coils of many turns of heavy copper tubing ending in the valve A at the bottom of a cylindrical container. When this valve is opened, the compressed air rushes out cooling itself and pouring up through the coils cooling the air that is still under pressure inside them. After this process has been going on for 10 or 15 min., the high-pressure air is already so cold before it reaches the valve A that the further cooling caused by the expansion is sufficient to liquefy it. The liquid air collects in a pool above B and is drawn off from time to time through the valve B. There is always

enough cold air not liquefied or evaporating from the pool at *B* to keep the coils in the liquefier precooled. The liquid air is drawn off into vacuum-jacketed containers like thermos bottles and can be stored in these for periods of several days. The process of liquefying hydrogen is the same in principle though the lower temperature required and the danger of explosions

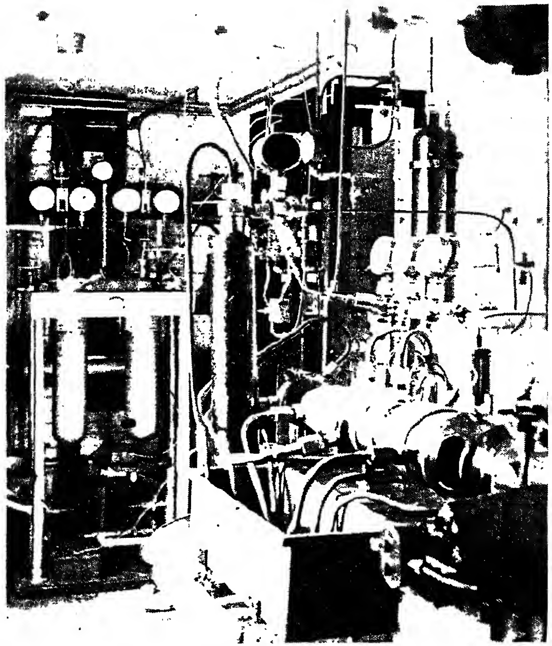


FIG. 78.—A photograph of the liquid-hydrogen plant in Palmer Laboratory.

introduce complications. That neither process is quite so simple as it sounds can be seen from Figs. 77 and 78 which are photographs of the liquid-air and liquid-hydrogen plants in Palmer Laboratory at Princeton.

SUMMARY

All substances can be solidified if the average motion of their molecules is reduced sufficiently by lowering the temperature. But there are always some molecules having sufficient kinetic

energy to break away from the rest so that even solids at low temperatures have some vapor pressure. In solids nearly all the molecules vibrate about fixed positions; in liquids, they move somewhat more freely but are still under the influence of attractive forces; in gases these intermolecular forces are dominated by the molecular motion. The vapor pressure of solids and liquids depends very much on the substance and on the temperature. The pressure of a mixture of vapors is the sum of the pressures that each would exert alone. Energy is required to convert a substance from the solid to the liquid state or from the liquid to the gaseous state. These energies are called the heat of fusion and the heat of vaporization, respectively.

If a solid and its liquid, a liquid and its vapor, or a solid and its vapor are in a closed container, they are said to be in equilibrium if the amounts of the substance in each state do not change. At a given temperature, the conditions for equilibrium are perfectly definite. For example, liquid will evaporate until the space above it is filled with vapor at a pressure characteristic of the liquid and the temperature. Graphs showing these relations are called equilibrium diagrams. Generally only two states, liquid and solid, gas and solid, or gas and liquid can exist at any particular temperature and pressure, but at certain conditions called triple points all three exist together in equilibrium. Every substance has a critical temperature above which it exists only as a gas.

The freezing point of a liquid is the temperature at which solid and liquid are in equilibrium under a given pressure. The freezing point varies somewhat but not rapidly with the pressure for most substances. The boiling point is the temperature at which the vapor pressure of a liquid becomes equal to the pressure of the atmosphere on it.

Air and other gases are liquefied by simultaneous cooling and compression. The last stage in the cooling is achieved by allowing the gas itself to expand suddenly.

ILLUSTRATIVE PROBLEMS

1. Two hundred grams of steam at 200°C. are passed into a 300-g. copper calorimeter containing 0.5 kg. of water and 0.5 kg. of ice at 0°C. What is the final temperature of the mixture?

In this type of problem, there is uncertainty from the statement of the problem as to the final state of the mixture. It is possible to have all ice,

ice and water, all water, water and steam, or all steam in the calorimeter. The general method of procedure is to reduce the contents of the calorimeter to a transition temperature, in this case either 0 or 100°C., and then see whether there is an excess of heat or lack of it left over and apply this heat to the entire contents to determine the final temperature. In this problem, we shall raise the calorimeter to 100°C. and reduce the contents to water at that temperature.

The water equivalent of the calorimeter is 300 g. × 0.0931 = 28 g., using the specific heat of copper from Table 7, page 156. To melt the ice at 0°C. requires

$$0.5 \text{ kg.} \times 1{,}000 \text{ g./kg.} \times 80 \text{ cal./g.} = 40{,}000 \text{ cal.}$$

We now have 500 g. + 500 g. + 28 g. = 1,028 g. of water at 0°C. To heat this to 100°C. requires

$$1{,}028 \text{ g.} \times 1 \text{ cal./g. °C.} \times 100\text{°C.} = 102{,}800 \text{ cal.}$$

so that the total heat absorbed by the ice, water, and calorimeter in rising to 100°C. is

$$102{,}800 \text{ cal.} + 40{,}000 \text{ cal.} = 142{,}800 \text{ cal.}$$

The heat given off by the steam in dropping to 100°C. is

$$200 \text{ g.} \times 0.467 \text{ cal./g. °C.}(200\text{°C.} - 100\text{°C.}) = 9{,}340 \text{ cal.},$$

where 0.467, the specific heat of steam at constant pressure, is assumed not to vary with temperature. The heat lost by the steam in condensing at 100°C. is

$$200 \text{ g.} \times 540 \text{ cal./g.} = 108{,}000 \text{ cal.}$$

so that the total heat lost by the steam is

$$108{,}000 \text{ cal.} + 9{,}340 \text{ cal.} = 117{,}340 \text{ cal.}$$

Therefore, the heat available at 100°C. to vaporize water in the calorimeter is

$$117{,}340 \text{ cal.} - 142{,}800 \text{ cal.} = -25{,}460 \text{ cal.}$$

The negative sign shows that the heat available cannot raise the system to so high a temperature as 100°C. The final temperature is given by

$$25{,}460 \text{ cal.} = 1{,}228 \text{ g.} \times 1 \text{ cal./g.°C.} \times (100\text{°C.} - t)$$

$$t = \frac{122{,}800 - 25{,}460}{1{,}228 \ 1/\text{°C.}} = 79.3\text{°C.}$$

Therefore, the calorimeter contains 1.2 kg. of water at 79.3°C.

PROBLEMS

1. Fit the curve of Fig. 69 with a power series including quadratic terms, and plot the curve so obtained. Does this formula fit Fig. 68?

2. Fit the curve of Fig. 68 with a formula of the form $\log p = A + B/T$ where T is the absolute temperature. This formula was proposed by Young in 1820. It can be improved by adding a term C/T^2.

3. How many calories are required to convert 100 g. of ice at −10°C. into steam at 100°C.? If the heat of combustion of illuminating gas is 136,000 cal./cu. ft., how much gas would be required for this purpose?

4. Five hundred grams of ice at −5°C. are placed in 1 l. of water at 50°C. Describe the final state of the mixture.

5. Seven kilograms of iron at 80°C. is dropped into a 200-g. copper calorimeter containing 0.5 kg. of ice at −10°C. What is the final temperature of the mixture?

6. The energy of a lead bullet in striking a steel target is largely converted into heat. If all the heat produced were used in warming the bullet, what must have been its velocity in order to be melted by the impact? Assume a temperature of 90°C. on striking the target.

7. From what height must (*a*) a block of ice and (*b*) a lump of tin have fallen (in a vacuum) in order to be melted on striking the ground? Assume all the energy of the falling object to be used in the melting process.

8. The density of water at 100°C. is 0.958 g./cc., whereas that of steam at the same temperature is 0.000589 g./cc. If it takes 539 cal. to convert each gram of water into steam, what fraction is retained in the form of heat by the steam and what fraction is spent in expansion?

9. In Chap. V, Prob. 18, if the excess energy of the bullet divides equally between the block and bullet, and if they are originally at 27°C., what will be the final temperature of each? Specific heat of wood is 0.42. Heat of fusion of lead 5.86 cal./g. at 327°C.

10. How much hydrogen must be burned in order to melt 1 kg. of iron originally at 30°C.? The heat of fusion of iron is about 7.89 cal./g. at 1530°C.

11. A certain waterfall is 100 m. high. The water above the fall has a temperature of 10°C. Before striking the bottom, 5 per cent of the water evaporates. What is the temperature of the water flowing away from the bottom of the fall if no heat is lost by conduction to the air?

12. Compute the difference between the heat emitted and absorbed going around the cycle of Par. 14 due to change of temperature and change of state. To what is this difference due? Specific heat of ice 0.50. Heat of sublimation of ice at −10°C. 732.9 cal./g. Heat of fusion of ice 79.7 cal./g.

CHAPTER X

STATIC ELECTRICITY

1. There is only one good thing that can be said for the summer climate of Princeton, New Jersey: very often, at the end of a sultry July day, nature puts on an electrical display that is really magnificent. In the course of half an hour or so, thousands of kilowatt hours of energy are blown off into lightning flashes that split the heavens. Sparks hundreds of feet long jump from cloud to cloud or from cloud to earth, making the finest efforts of laboratories or power stations seem puny. Two hundred or even one hundred years ago many well-educated people lived their lives through in civilized communities without seeing any other electrical phenomena than such thunderstorms as these. But today the community is primitive indeed that cannot boast of a radio, a lighting plant, or at least a flashlight. We live in a world of transmission lines, telephones, vacuum cleaners, and other electrical gadgets. Yet it may be doubted whether most people know much more about electricity than their forefathers did, when it comes to an understanding of its properties and the laws that govern its behavior. In attempting to study electricity, we shall not start with the most familiar phenomena, such as electric lights or thunderstorms, but with some very simple experiments that have been familiar to a few people for a very long time, the experiments on frictional electricity. These experiments are not very impressive or spectacular, yet they form the basis for the theory of electricity whose application has changed man's mode of life so greatly in the past hundred years.

Frictional Electricity

2. On a dry day, the reader himself can easily perform the first of these experiments on frictional electricity. Let him take a piece of hard rubber such as a black fountain pen and rub it vigorously on a woolen coat sleeve. It will be found to have acquired a new property. It will pick up small bits of paper

or dust. In time, it loses this property, particularly rapidly if the weather is not extremely dry. This same experiment can be done with many other materials. Amber, for example, was known by the Greeks to become attractive when rubbed and has given its Greek name to the subject of electricity. When a body has acquired this attractive property as the result of rubbing, it is said to be electrified. We might give a long list of bodies that have been found to become electrified when rubbed together; instead, we shall study a few examples in detail. Apparently electrified bodies exert forces on neighboring bodies but evidently these forces are small since large heavy objects are not moved. To study them, therefore, we want to reduce resistance to motion as much as possible. This may be done by making the inertia of the moving bodies as small as possible and by mounting them so that the frictional or gravitational resistance to motion is as small as possible.

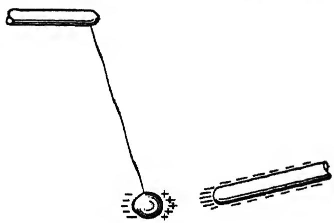


FIG. 79.—Pith ball attracted by a charged rubber rod. Induced charges on the ball are equal in magnitude and opposite in sign. Unlike charges attract each other.

Suppose that instead of bits of dust or paper we use as a test body a small sphere made of pith coated with aluminum and hung by a thread from a support. Then we know that a very small force will be needed to deflect this pith-ball pendulum. Instead of a fountain pen and a more or less woolen sleeve, we shall use a solid rod of hard rubber and a cat's fur. Rub the rod vigorously with the cat's fur, and bring it up toward the pith ball. The pith ball is clearly attracted to the rod. Bring up the cat's fur instead of the rod. The pith ball is again attracted. Now use the rod again but let the pith ball touch it. As soon as contact has been made, the ball is strongly repelled. Apparently, electric forces may be either attractive or repulsive. Let us investigate further. Take a second rubber rod, and arrange a cradle suspended by threads which will hold one of the rods in such a way that it can turn easily (see Fig. 80). Rub both rods with the fur, put one of them in the cradle, and bring the other up close to one end of it. The suspended rod will be repelled, turning in the horizontal plane. Bring up the cat's fur instead of the second rod; the suspended rod will now be attracted instead of repelled. Repeat the experi-

ment, using glass rods and a piece of silk. The same effect will be observed though it will not be so marked. In each case, we have seen that identical bodies, electrified in the same way, repel each other. Now rub a rubber rod with the cat's fur, and put it in the cradle but bring up to it one of the glass rods that has been electrified with the silk. The two rods attract each other. Apparently, the glass rod has the same kind of electrification as the cat's fur. Many repetitions of experiments of this sort with a variety of substances have shown that there are two and only two kinds of electrification and that the kind of electrification which a substance will acquire depends on the substance with which it is rubbed. For example, if the glass had been rubbed with fur, it would have repelled the rubber rod and we would say that it had the same kind of electrification as the rubber.

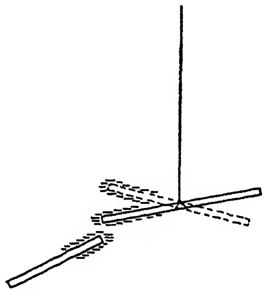


FIG. 80.—Repulsion between two negatively charged rubber rods. Like charges repel each other.

3. We shall not stop to outline the steps by which we have arrived at our present explanation of these experiments, interesting though these steps are from the historical and philosophical point of view. But we do want to remind the student that the statements we are about to make are assumptions. They are assumptions that have proved useful in correlating many observations and to that extent the theory of which they are a part may be called true. According to our present views, all matter is electrical in nature. In every atom of every element, there are always two kinds of electricity present which we call positive electricity and negative electricity. We can think of every atom as made up of a number of little bits of positive electricity and little bits of negative electricity. These positive and negative electric charges, as we call them, attract each other and, when present in equal numbers, nullify each other's effect or, in technical language, neutralize each other. Ordinarily, therefore, a body contains equal numbers of positive and negative charges which completely neutralize each other so that the body shows no electrical properties. But if the surfaces of two bodies are rubbed together, some of the negative charges may be scraped

off one surface and stick to the other. The result is that the first surface is left with more positive charges than negative, *i.e.*, acquires a resultant positive charge, whereas the second surface has sticking to it a lot of extra negative charges, *i.e.*, has acquired a resultant negative charge. In the case of the fur and rubber, the negative charges are scraped off the fur and stick to the rubber giving it a negative charge. With the glass and silk, the negatives come off the glass onto the silk, leaving an excess of positive charge behind. It is possible to arrange substances in a list such that each substance will be positively charged if rubbed by a substance lower on the list and negatively charged if rubbed by a substance higher on the list. A brief list of this sort is as follows:

Fur
Wool
Glass
Silk
Hard rubber
Sealing wax
Resin
Sulfur.

We have established that there are two kinds of electric charge which we call positive and negative, and that unlike charges attract each other whereas like charges repel each other. We have also suggested that these charges, at least the negative ones, can be transferred from one body to another by the mechanical process of rubbing. We now want to consider this question of transfer of charge more fully.

Conductors and Nonconductors

4. Since like charges repel each other, there is always a tendency for the charge on a body to spread out as much as possible, for the little particles of charge to get as far away from each other as possible. But this cannot happen unless the charges can move around in the body. It is found that the possibility of this motion depends very much on the nature of the substance. In the materials in the list given in the last paragraph, such motion hardly occurs at all so that the charges produced by rubbing stay where they are left by the frictional forces.

In metals, on the other hand, the charges move about with comparative ease, and for this reason it was thought for a long time that metals could not be electrified. What happened was that the charges produced by friction continually leaked away before their presence could be detected. Substances through which charges can move fairly easily are called *conductors*, while those through which little or no motion can occur are called nonconductors or *insulators*. There is no sharp line that can be drawn between the two classes. In general, metals are good conductors and nonmetals are insulators, but some alloys are pretty poor conductors, and many nonmetals do allow the motion of electric charges to some extent.

5. Suppose we consider some experiments that illustrate this difference between conductors and insulators. Use the pith

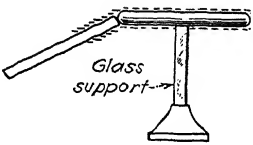


FIG. 81*a*.—Communication of a negative charge to a conductor.

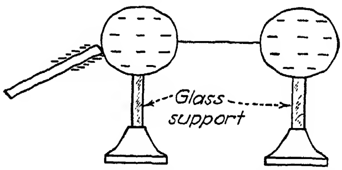


FIG. 81*b*.—Negative charge carried from one sphere to another by a wire.

ball again as a test of the presence of electrification. Take a long cylinder of brass on an insulating support such as shown in Fig. 81*a*. Electrify a rubber rod as usual, and then rub the rod over one end of the brass cylinder, giving the negative charges a chance to get onto the brass. The end of the brass is now charged as is shown by its attraction for a pith ball. Try the pith ball on the other end of the brass. It is charged too. Some of the charge has moved to the other end of the cylinder. Repeat the experiment, using a cylinder of glass. Very little charge will get onto the end that is touched with the rubber and none will flow to the other end. Repeat the experiment, using two insulated brass balls instead of the cylinder and connecting them first with a piece of wire and then with a silk thread. The wire will transmit the charge, the thread will not.

Nature of Moving Charges

6. Until twenty or thirty years ago, positive and negative electric charges were thought of as moving like fluids from place to

place and there was no reason to think that one was more mobile than the other. We believe now that both positive and negative charges are carried by small particles but that the negative particles, called electrons, are much more mobile than the positive charges. In fact, they are the only charges that move about in solid materials. But the motion of electrons away from some point leaves a resultant positive charge there and gives an effect which is indistinguishable from the effect that would be produced by the flow of positive charge to that same point. In other words, the flow of negative charge in one direction is exactly equivalent to a flow of positive charge in the other direction. In the next few chapters, we shall have much to say about the nature and behavior of electrons. For the present, they may be thought of as extremely tiny negative charges of mass several thousand times less than that of an atom and able to move about freely in conducting materials.

The Electroscope

7. The experiments we have been describing are of a simple qualitative type. Before much can be known about electricity, they must be made quantitative and measuring instruments must be devised. The first of these which we now want to discuss is an instrument known as an electroscope. It depends very simply on the principles that have been presented and will be used repeatedly in our study of electrical phenomena. In one of its simplest forms, it consists of a vertical brass rod about 6 in. long to which a light metallic vane of gold leaf is attached by a hinge about 2 in. from the bottom. For convenience, a flat disk of brass is screwed to the top of the rod. This is called the table of the electroscope. The rod runs through a block of amber or other insulating material as a support, and the insulating block is mounted in the top of a glass-sided case which protects the hinged vane from air currents but allows the motion of the vane to be observed. Now suppose that a charged rubber rod is brought into contact with the brass disk of the electroscope. Some of the electrons on the rubber will go off to the brass. Once there, they will try to get as far away from each other as

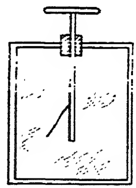


FIG. 82.—An electroscope. A gold leaf is attached to an insulated metal rod.

possible since they repel each other. This means that some of them will flow down into the rod and into the vane so that the vane and the rod will both be negatively charged. Consequently, the rod will repel the vane which will stand out at an angle to it. Since the moment of force tending to bring the vane down against the rod depends on the angle and on the weight of the vane, the deflection of the vane will be greater the greater the electric force of repulsion and a light vane will be required if small electrical forces are to be studied. For a given instrument, the electrical force of repulsion will depend on the charge given to the instrument and therefore the angle at which the vane stands is a measure of the charge on the instrument. Such an instrument frequently has a scale against which the deflection of the vane can be read.

Induced Charges

8. In Par. 3, we mentioned that all matter is supposed to be electrical and always to contain both positive and negative charges. Most of the time, positive and negative charges are present in equal numbers so that they completely neutralize each other and the body containing them shows no electrical properties. Remembering this, consider what happens in a conducting body if an electrical charge is brought near. Take the insulated brass cylinder of Fig. 83, for example. Suppose a charged rubber rod is brought near one end of it without touching it. The negative electrons in the brass will be repelled by the negative electrons on the rubber. Since brass is a conductor, they can move under this repulsive force and some of them will go to the far end of the brass cylinder, getting as far away as possible and giving a negative resultant charge to that end of the brass. Their departure from the near end of the brass will leave a dearth of electrons there and produce a resultant positive charge. Electrons will move away from the negative charge on the rubber until the attraction of the positive charge built up by their departure is big enough to neutralize the repulsion of the negative charge on the rubber. Thus we have charged the two

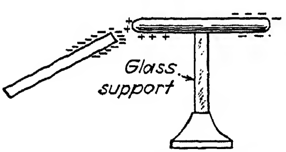


FIG. 83.—Positive and negative charges induced in a conductor by a neighboring negative charge.

ends of the brass cylinder without even touching it. Charging by this process is called induction, and the charges produced in this way are called induced charges. In the case just described, the induced charges will flow together again as soon as the inducing charge of the rubber is removed. But suppose we do the experiment with two insulated brass spheres connected by a wire. The process will be the same. One sphere will become negatively charged and the other positively charged. But now take away the connecting wire before taking away the inducing charge. The spheres are now insulated so that when the inducing charge is taken away the induced charges cannot flow together. The spheres are permanently charged. Moreover, there has been no change of the total charge on the spheres, merely a redistribution. Consequently, the positive charge on one must be just as big as the negative charge on the other. This is a general property of induced charges. For every positive charge induced in one place, there must be an equal negative charge induced somewhere else. In doing electrical experiments, charging by induction is usually much more convenient than charging by direct contact. As an example, we shall consider in detail the charging of an electroscope by induction.

Charging an Electroscope by Induction

9. Suppose that we have the same electroscope as in Par. 7 but that we bring up the negatively charged rubber rod close to the table of the electroscope without making actual contact. Then some of the electrons in the table are repelled to the lower end of the electroscope and the vane is deflected by the negative charges. If the rod is taken away, the charges flow together again and the vane sinks back to its vertical undeflected position. Now suppose that before taking away the inducing charge on the rubber rod we connect the electroscope to the earth by touching its table with a finger. Since the earth now offers a refuge more remote from the negative inducing charge than the lower end of the electroscope, electrons from the table will flow away to the earth and the electrons that have previously gone to the vane will flow back to take their place, leaving little or no charge on the vane. Its deflection will diminish. In effect, the table of the electroscope is one sphere of our experiment of Par. 8 and the earth is the other. It makes no difference at what point on the

electroscope the connection to earth is made. Before removing the inducing charge, take the finger off the table of the electroscope, breaking the connection to earth. An excess positive charge is left on the electroscope but it is concentrated on the table where the electrons have been driven away by the negative inducing charge on the rubber rod. Now remove the rubber rod. The reduced number of electrons now distributes itself, leaving a resultant positive charge everywhere on the electroscope and therefore producing a deflection of the vane. This then is a way of charging an electroscope positively by induction. In practice, the whole process is carried out simply and easily in a few seconds. To charge the electroscope negatively, an exactly similar process can be carried out, using the positive charge on a glass rod as the inducing agent.

Coulomb's Law and the Unit of Charge

10. In Chap. IV, we discussed the law of gravitational attraction and stated that the force of attraction between two bodies was proportional to the product of their masses and inversely proportional to the square of the distance between them, or mathematically, $F = Gm_1m_2/d^2$ where G was a constant of proportion. We also described the Cavendish experiment with a torsion balance as an experimental proof of this law. A French contemporary of Cavendish, Coulomb, used a similar apparatus to find out exactly what the law of force between electric charges was. A schematic sketch of his apparatus is given in Fig. 84. AB is a stiff light rod of some insulating material (actually a piece of thread stiffened with "Spanish wax"). At A is a small sphere of conducting material; at B is a counterweight. The rod AB swings freely in a horizontal plane at the end of a silver wire S. The upper end of S is carried in a torsion head that can be turned through an angle which can be read on a scale. From previous experiments, it was known how much force was necessary to twist the suspension through any given angle. To perform the experiment, the torsion

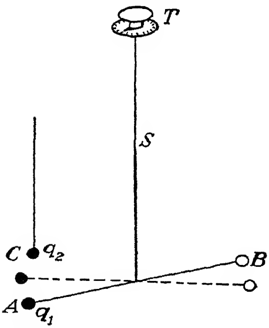


FIG. 84.—Coulomb's experiment. The fixed charge q_2 at C repels the charge q_1 to A thus twisting the wire S.

head was set so that the ball at A was almost in contact with a similar fixed ball at C. Both balls were then given like charges. The movable ball then moved away until the twisting force in S was just big enough to counteract the force of repulsion between A and C. The angle through which AB had turned was read on a scale on the outside glass case of the instrument (not shown) and the torsional restoring force calculated. The torsion head was then twisted until the distance between A and C was halved and the torsional force again computed. It was found to be four times as great as before. Series of experiments of this kind and further experiments in which the amount of charge on the balls was varied led Coulomb to the conclusion that:

The force between two electric charges is proportional to the product of the charges and inversely proportional to the square of the distance between them. It is a force of attraction if the forces are unlike, a force of repulsion if the charges are alike.

11. Later experiments show that the force also depends on the nature of the material between the two charges. Mathematically, then, we have the following law:

$$F \propto \frac{q_1 q_2}{d^2} \qquad \text{or} \qquad F = \frac{B q_1 q_2}{d^2} \tag{1}$$

where F is the force, attractive if negative, repulsive if positive, q_1 and q_2 are the charges, d is the distance between them, and B is a constant of proportionality whose magnitude depends on the units in which the other quantities are measured and on the nature of the medium separating the charges. B is like G in the gravitational equation. But here we have two kinds of charge, and the sign of the force depends on the sign of their product, whereas we know only one kind of mass, and the gravitational force is always attractive. Another difference emerges when we consider how the charges are to be expressed quantitatively. Before we discussed the law of gravitation, we had defined a unit of mass, the standard kilogram, and a unit of force in terms of the acceleration imparted to a unit of mass. Consequently G was determined experimentally in terms of these units. But we have not yet defined any unit of electric charge; therefore we cannot determine B.

12. We have pointed out that we now know that all electric charges are built up of tiny identical charges. We might well

say that they are natural units of charge. If we could put one of these electrons on each of the balls A and C in Coulomb's experiment, we might find a value of B appropriate for air. Unfortunately this is an impossible experiment, and the electronic charge is much too small to be a convenient unit, however natural it may be. To avoid future confusion, we hasten to introduce the coulomb, the so-called practical unit of electric charge. In terms of electrostatics, it has no logical basis, and we must wait until we have introduced the ampere and the volt to explain its origin and give it a logical definition. But we can define it arbitrarily in terms of Coulomb's law as follows:

13. A charge of one coulomb repels an equal like charge in vacuum at a distance of one meter with a force of 9.0×10^9 newtons or, putting it in terms of the general law for the force between two charges,

$$F = 9.0 \times 10^9 \frac{q_1 q_2}{d^2} \text{ newtons} \tag{2}$$

if the charges are in coulombs, the distance in meters, and the space between the charges is entirely empty. Evidently we have defined the coulomb by arbitrarily giving a certain value to the constant of proportion B, of Eq. (1). To be consistent with the later developments of the subject, this equation is usually written in the form

$$F = \frac{q_1 q_2}{4\pi\epsilon d^2} \tag{3}$$

where ϵ is called the dielectric constant and depends on the nature of the medium separating the charges as well as on the units used for force, distance, and charge. The dielectric constant for a vacuum is called ϵ_0 and in our system of units has a value of $\dfrac{1}{4\pi \times 9 \times 10^9} = 8.84 \times 10^{-12}$ as may be seen from Eqs. (2) and (3).* The value for air is very nearly the same, and the difference can be neglected for all but the most exact calculations.

14. We have seen that it is convenient in dealing with masses and lengths to introduce secondary units differing by powers of

* From Eq. (3) we can see that the units in which ϵ is expressed are complex. They are coulomb²/newton × m.² Similarly B is expressed in newton × m.²/coulomb.² These are equivalent to farad/m. and m./farad, respectively, where the farad is the unit of electrical capacity discussed in Chap. XI, Par. 16.

10 from the fundamental unit and to give them self-explanatory names such as kilometer or millimeter. Thus we may consistently call a charge of one one-thousandth of a coulomb a millicoulomb or a charge of one one-millionth of a coulomb a microcoulomb. Expressed in powers of 10, we have

$$1 \text{ coulomb} = 10^3 \text{ millicoulombs} = 10^6 \text{ microcoulombs.}$$

15. In the electrostatic experiments we have been describing in this chapter, the forces observed were evidently very small. Only small masses like bits of paper or pith balls could be moved about readily by electrostatic attractions or repulsions. Heavier bodies like glass and rubber rods had to be hung in cradles where little force was needed to turn them before they would respond to the forces caused by frictional electricity. Comparing these observations with the large numerical coefficient of Eq. (2), we see that the number of coulombs in the charges produced by frictional effects must be very small. This depends, of course, on the particular conditions, the size and shape of the rubber rod used for example, the humidity of the atmosphere at the time, and so on. Roughly speaking, a rubber rod 1 ft. long (0.3 m.) and 1 in. in diameter (25 mm.) rubbed by an average cat's fur on an average day will acquire a negative charge of about one-tenth of a microcoulomb, or 1.0×10^{-7} coulomb.

16. It is evident from the preceding paragraph that we would not choose the coulomb as the unit of charge if we were expecting to confine our study of electricity to frictional effects. As a matter of fact, there is another unit which is both more convenient and more logical in this division of the subject. Nevertheless, we mention it with hesitation and will not use it at all because we wish to keep to the "practical" system throughout. This other unit is the electrostatic unit of charge sometimes called the statcoulomb. It is defined by Eq. (1) if the force F is expressed in dynes, the distance in centimeters, and the coefficient B is taken as one. The unit of charge so defined forms the basis of a whole system of electrical units called the electrostatic absolute system and is used for parts of the treatment of electricity in most books. One such unit, *i.e.*, 1 statcoulomb, is equal to $10^{-9}/3$ coulomb.

Electrostatic Generators

17. The generation of electric charge by the personal application of a cat's fur to a piece of rubber is obviously not the method used by the great public-utility companies of today to light cities and run subways. As you know, the methods they

use depend on the interplay of electric and magnetic forces, a subject that we are not yet ready to discuss. But there is a type of electric generator which depends on the kinds of effects that we have been talking about. The different forms of this kind of machine used in the past are variously known as induction machines, influence machines, Wimshurst machines, etc., according to the particular type. For practical purposes, they are as dead as the dodo though they are still useful occasionally for lecture purposes. But there has been developed in recent years an electrostatic generator that has already proved valuable for certain special purposes and which may well grow to be used fairly extensively. The first generator of this type was built by its inventor Van de Graaff at Princeton in 1930 and is still in working order. It will be described presently, but first we need to mention two points about the behavior of charges that we have not yet discussed.

18. The first of these points is often discussed at length in elementary texts with a wealth of experimental detail to support the conclusion. This seems hardly necessary, and we shall merely state the conclusion and show that it is reasonable. It is that the charge on a conducting sphere is all on the outside and that there is no electrical force on the inside of the sphere. Since there is a repulsive force between like charges, they will naturally try to get as far away as possible from each other. On a sphere, this means getting as far away from the center of the sphere as possible since this will make the average distance from any one charge to all the others as great as possible. Consequently, any charge that is put on a spherical conductor, whether solid or hollow, distributes itself over the surface. To show that the resultant force of this surface charge on a charge introduced anywhere inside the sphere is zero is a little beyond the scope of this course. But it may be of interest to mention that the proof of this fact depends on Coulomb's inverse-square law of force between charges. In fact, Cavendish established the inverse-square law by proving experimentally that there was no charge on the inside of a closed conductor.

19. The second question to be cleared away before going to the Van de Graaff generator is that of the discharge of electricity through air. In the simple experiments we have described, we have tacitly assumed air to be a perfect insulator. This is not

strictly true. If a conductor is very highly charged, some of the charge leaks off into the air either gradually or in a sudden spark. The degree to which this happens depends not only on the voltage to which the conductor is charged but on the shape of the conductor. The leakage is much greater from points or sharp edges on a conductor than from the smooth flat surfaces. Speaking in terms of a crude analogy, the electrons can evaporate much more easily from a sharp point where they are almost surrounded by air than from a flat surface from which there is only one direction of possible escape, just as a small droplet of water has a higher vapor pressure than a flat surface. Consequently, if good insulation is wanted, everything is made as smooth and flat as possible, whereas if we want to discharge a conductor into the air, we put a sharp point or points on it.

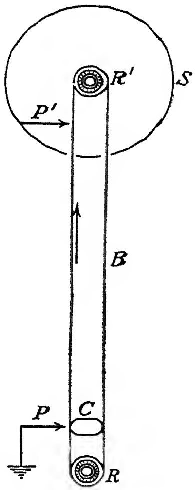


FIG. 85.—Original Van de Graaff machine (schematic drawing). Electrons are sprayed on the belt B at P and removed at P' where they spread over the sphere S.

20. A schematic and somewhat simplified version of a Van de Graaff generator is shown in Fig. 85. S is a large hollow copper sphere mounted on a heavy pyrex glass rod (not shown) so as to be insulated. B is a belt made of nonconducting silk ribbon and running on two rollers R and R'. The lower roller R is mounted on the shaft of a small motor. A charge is built up on the sphere S by having the belt continually charged at P and continually discharged at P'. One way of doing this is to have the points at P connected to an auxiliary device giving a fairly high negative potential—say 10,000 volts. Then the negative charge on the points P induces a positive charge on the conductor C behind the belt. This helps to draw electrons off the points toward the belt which intercepts them and carries them up toward the sphere. No matter how much charge has already gone to the sphere, none is left inside and there is no resultant force on the belt charge once it gets inside the sphere (see Par. 18). But when the negative charge on the belt gets opposite the points P', it induces positive charges in the points which draw electrons from the belt, reducing its charge. The points P' are connected to the sphere so that its negative

charge increases. This process continues until the sphere is charged to so high a voltage that as much charge leaks away from it through the support and the surrounding air as is brought to it by the belt. In practice, there is a somewhat more complicated arrangement of points and conductors than has been described

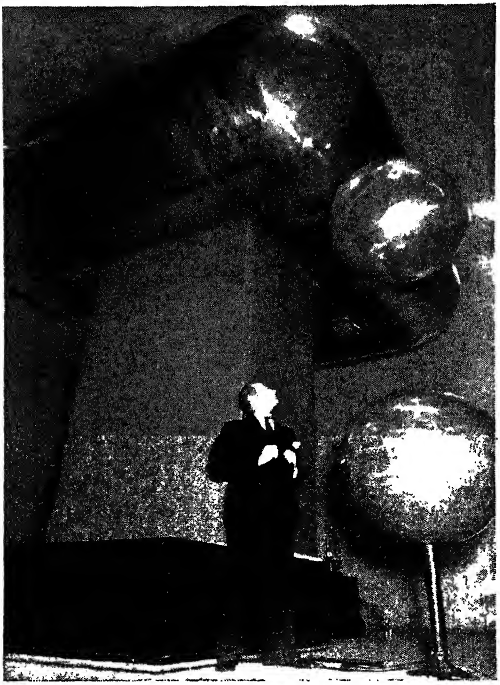


FIG. 86.—Van de Graaff generator used for x-ray cancer treatment at the Huntington Memorial Hospital in Boston. The charging belts run up inside the rectangular insulating column on which the high voltage terminal rests. The x-ray tube is behind this column. (*Courtesy of Dr. J. G. Trump.*)

here, but their object is merely to double the efficiency of the machine by making the belt carry a charge of the opposite sign (a positive charge in our description) away from the sphere on its descent. The principle is the same.

21. It may be wondered why so simple a device is not in general use. The answer is in the smallness of the current that

the belt carries. It can be measured by connecting the copper sphere to ground through a galvanometer and is found to be of the order of a few microcoulombs per second. The great advantage of the device is that it builds up very high voltages. For example, each of the two original generators constructed by Van de Graaff will build up a potential of some two hundred thousand volts so that if we make one positive and one negative we get a difference of potential of about four hundred thousand volts. To do this by electromagnetic means would require an enormous amount of expensive equipment. The importance of getting such high voltages as these even with comparatively low power will be evident later on when we come to the discussion of nuclear physics or "atom smashing" as the newspapers call it.

SUMMARY

Rubbing two different substances together may give them the power of attracting light pieces of paper or dust. Materials in this condition are said to be electrified. There are two kinds of electrification that are interpreted as caused by the presence of two kinds of electric charge, positive and negative. The force between two charges is proportional to the product of the charges and inversely proportional to the square of the distance between the charges. It is a repulsion if the charges are alike, an attraction if they are unlike. The unit of charge is the coulomb. If we use this unit and m.k.s. units for the other quantities, the mathematical expression for the force between two charges in vacuum is

$$F = 9 \times 10^9 \frac{q_1 q_2}{d^2}.$$

This expression is nearly correct if the charges are in air.

There are always both positive and negative charges present everywhere. Electrification or "charging" is a condition where one kind of charge is present in excess of the other. Negative charges consist of electrons and are more mobile than positive. In conductors, they can move freely; in insulators, with difficulty or not at all.

Charges in conductors can be moved by other charges in the neighborhood. A positive charge near one of two connected conducting spheres will draw a negative charge to that sphere and force a positive charge to the second sphere. This process is

called charging by induction. The charges on the spheres can be isolated if the connection between them is broken while the inducing charge is still present.

An electroscope consists of a light conducting vane hinged to a vertical metallic rod supported by an insulator. When the electroscope is charged, the vane is repelled by the rod, standing out at an angle to it. An electroscope is usually charged by induction.

Many types of electrostatic generators of electric power have been built, but the only one of present importance is that invented by Van de Graaff. In it, charges are carried to the interior of a metallic sphere by a silk belt. The charges flow out to the surface of the sphere and build up on it until the insulation begins to leak. Potentials of several million volts are obtained with these generators, but the currents are very minute.

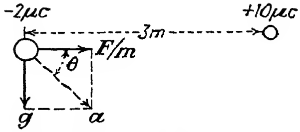


FIG. 87.—Accelerations of an electrostatic charge under the influence of another charge and gravity.

ILLUSTRATIVE PROBLEMS

1. A 1-g. pith ball carrying a charge of -2 microcoulombs is placed 3 m. horizontally from a charge of 10 microcoulombs. If the pith ball is released, what is its initial resultant acceleration?

The acceleration will be the resultant of a horizontal component F/m due to the electrostatic force and a vertical component g due to gravity, Fig. 87. The electrostatic force is given by Eq. (1), page 217

$$\begin{aligned} F &= B\frac{q_1q_2}{d^2} \\ &= 9 \times 10^9 \frac{\text{newton-m.}^2}{\text{coulomb}^2} \times \frac{-2 \times 10^{-6}\text{ coulomb} \times 10^{-5}\text{ coulomb}}{(3\text{ m.})^2} \\ &= \frac{-18 \times 10^{-2}\text{ newton-m.}^2}{9\text{ m.}^2} = -0.02\text{ newton.} \end{aligned}$$

The minus sign indicates attraction toward the 10-microcoulomb charge The acceleration, which is horizontal, produced by this force is $a = F/m$

$$a = \frac{0.02\text{ newton}}{1 \times 10^{-3}\text{ kg.}} = 20\text{ m./sec.}^2$$

The resultant acceleration is obtained from the Pythagorean theorem.

$$\begin{aligned} a^2 &= (20\text{ m./sec.}^2)^2 + (9.8\text{ m./sec.}^2)^2 \\ a^2 &= 496(\text{m./sec.}^2)^2 \\ a &= 22.3\text{ m./sec.}^2 \end{aligned}$$

The direction of a from Fig. 87 is given by

$$\tan\theta = \frac{9.8 \text{ m./sec.}^2}{20 \text{ m./sec.}^2} = 0.49$$

$$\theta = \arctan 0.49 = 26°.$$

PROBLEMS

1. A charge of 30 microcoulombs is 3 m. from a charge of −20 microcoulombs. What is the force on the first charge?

2. Two equal and opposite charges of 10 microcoulombs are placed 6 cm. apart. Find the force on unit charge placed (*a*) on the line joining the charges 3 cm. from the positive charge, (*b*) half way between the charges, (*c*) on the line joining the charges 3 cm. from the negative charge.

3. Two equal and opposite charges of 5 microcoulombs are placed 6 cm. apart. Find the force on unit charge placed at a point on the perpendicular bisector of the line joining the charges 4 cm. away from the line.

4. A charge of +4 microcoulombs is placed on an 80-mg. pith ball situated 30 m. directly above a charge of −5 microcoulombs. What is the total force acting on the pith ball and what is its initial acceleration?

5. A charge of 0.2 microcoulomb is placed on a 50-mg. pith ball situated 3 m. directly above a charge of 4 microcoulombs. What is the total force acting on the pith ball and what is its initial acceleration?

6. A 100-mg. pith ball carrying a charge of −1 microcoulomb is placed 3 m. horizontally from a charge of 1 microcoulomb. If the pith ball is released, what is its initial resultant acceleration?

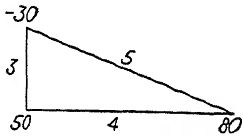


FIG. 88.—A triangle 3, 4, and 5 m. on a side with electrostatic charges in microcoulombs as shown at its vertices. (Prob. 7.)

7. Three 50-mg. pith balls are placed at the vertices of a right triangle in a horizontal plane with sides as shown in Fig. 88. The balls are charged with 80, 50, and −30 microcoulombs. If the pith ball with −30 microcoulombs is released, what will be the horizontal magnitude and direction of its initial acceleration?

8. A tetrahedron 2 m. on a side has charges of −2 microcoulombs at two vertices of the base and 10 microcoulombs at the third vertex. If a 15-mg. pith ball with a charge of 0.2 microcoulomb is placed at the apex, what is the resultant force on it and its initial resultant acceleration?

9. Two identical pith balls, each of mass 10 mg., are suspended from the same point by threads 50 cm. long. What equal charges on the balls will cause them to separate until the angle between the supporting threads is 60°?

10. Two identical 50-mg. pith balls, each carrying a charge of 1 microcoulomb are suspended by two threads from the same point. If the angle between the threads is 90°, how long is each thread?

CHAPTER XI

ELECTRIC FIELDS AND POTENTIAL DIFFERENCES

1. Most intelligent people have at least some vague idea of what is meant by an electric current and are familiar with the terms "high voltage" and "low voltage" though they might be embarrassed if asked to explain the meaning of these expressions. But they would have very much less idea of what an electric charge is or of Coulomb's law. This is because the practical uses of electricity all depend on the motion of electric charges, not on their mere existence. But to understand the behavior of charges in motion, *i.e.*, electric currents, we need to consider more fully the forces that act on charges and to introduce some ideas which are most easily understood in connection with charges at rest. In the last chapter, we discussed Coulomb's inverse-square law of force between charges at rest. We saw that this was formally the same as the law of gravitational attraction between masses, and consequently some discussion of gravitation is to the point.

2. None of you can remember when you first became familiar with the law of gravitation (the fact, not the name). Perhaps when you first dropped a toy over the edge of your crib you hardly realized the cosmic significance of your experiment. But now you know that wherever you are on the earth's surface or above it there is always a gravitational force acting on you and on everything else around you. A region of this sort where there is always a force acting on a body that comes into the region is called by physicists a "field of force," and the force acting on some standard test body at any point is called the "field strength" or "field intensity" at that point. Thus the earth is surrounded by a field of gravitational force, and the strength or intensity of the field depends on the square of the distance from the center of the earth. The test body is a mass of 1 kg., and we would say that the strength of the gravitational field at the pole was 9.83 newtons/kg. and at the equator, 9.78 newtons/kg. If a strato-

sphere balloon flight could reach a height of 4,000 mi. instead of a mere 13 or 14 and a determination of the strength of the earth's gravitational field were made, a value of about 9.80/4 or 2.45 newtons/kg. would be obtained. If we traveled on out through space until the influence of the moon, the other planets, or the sun became comparable with that of the earth, not only the magnitude but the direction of the gravitational field would change. But wherever we were in space, the gravitational field would have a definite direction and magnitude which would express the resultant effect of the attractions of all the heavenly bodies near and far.

Electric Fields

3. Actually, our lives are confined to a small region on or near the earth's surface in which the strength and direction of the gravitational field are constant or so nearly so that we rarely have to concern ourselves with variations. But in the analogous electrical case, we do get great variations of the strength and direction of the electrical field in a small space. Moreover, we can control the field by moving charges around. To get an electrical field strictly analogous to the earth's gravitational field, we would have to have a very large sphere carrying a negative charge. Then a body carrying a unit positive charge and close to the surface of the sphere would everywhere be acted on by a force directed toward the center of the sphere, and the magnitude of the force would be approximately constant as long as the variation of the distance from the surface of the sphere was small compared with the radius of the sphere. Clearly these conditions were not approximated even in the simple electrostatic experiments described in Chap. X so that the problem of electric fields is, in practice, more complicated than that of gravitational fields. Before considering these problems, we should define precisely what we mean by the electric field strength, or electric intensity as it is often called.

The intensity of the electric field at any point in space is equal in magnitude and direction to the force per unit charge which would act on a positive charge placed at that point.

4. The electric intensity is a vector quantity since it requires both a magnitude and a direction for its complete specification. It is usually designated by E. From the foregoing definition, it

follows that the force F on a charge q at a point in space where the electric intensity is E is given by $F = qE$ just as the gravitational force W on a mass m in the earth's gravitational field is given by $W = mg$.

Field of a Point Charge

5. We saw that Coulomb proved the force between two charges q_1 and q_2 to be $q_1q_2/4\pi\epsilon d^2$ if appropriate units are used. Let us now apply this law to determine the electric intensity caused by a point charge. Let q_1 be a comparatively large charge on a small conducting sphere fixed in space. Let q_2 be a unit positive charge on some small movable body like a pith ball. Let d be the distance from the center of the sphere to the center of the pith ball. Then the force on the pith ball will measure the electric intensity at the point where it is situated. But Coulomb's law shows that this force will always be directed toward the center of the sphere (if q_1 is negative) and its magnitude will be given by $q_1/4\pi\epsilon d^2$.*

6. It may well be asked why it is desirable to introduce the notion of an electric field when Coulomb's law suffices to calculate the forces in question. The answer is that the electric field at any point in space may be the resultant of the effects of many different charges distributed in so complicated a way that a direct calculation of the forces is impossible. Yet the electric intensity still has a definite physical meaning at every point in space and may be directly or indirectly measured.

Lines of Force

7. Since the electric intensity caused by a charge has a definite direction at all points around the charge, it is possible to draw lines out from the charge such that their direction at any point represents the direction of the electric intensity at that point. Furthermore, it is possible to indicate the magnitude of the

* In studying gravitation (Chap. IV, Par. 8), we mentioned the proof that the resultant gravitational attraction of a sphere of uniform density was the same as if the mass were all concentrated at a point at the center of the sphere. A similar theorem shows that the effect of a charged sphere at points outside it is the same as if the whole charge were concentrated at the center of the sphere. Inside the sphere, the electric intensity is everywhere zero as we mentioned in Par. 18, Chap. X.

intensity by the number of lines drawn. Such lines are called lines of force. In Fig. 89, for example, we have the lines of force representing the field of force set up by a charge on a spherical conductor. If all the lines start from the charge, then the number going through any spherical surface surrounding the conductor is the same. But the area of such a surface increases with the square of its radius. Therefore, the number of lines of force cutting a square meter of such a surface must fall off as the square of the radius. But we have seen that the electric intensity also falls off as the square of the radius; therefore, the number of lines of force per unit area is proportional to the magnitude of the electric intensity. This is a perfectly general relation, and diagrams like Fig. 89 are of great value quite apart from any possible interpretation of lines of force as "stresses in the ether" or any attempt to use them in a strictly quantitative way. If the convention is made that lines of force always start on a positive charge and end on a negative charge, never starting or stopping in free space, then the lines of force will always crowd together in a region of great electric intensity and spread wide apart in a region of small electric intensity. In Fig. 90, the lines of force between two charged spheres are shown both for the case of like and for that of opposite charges.

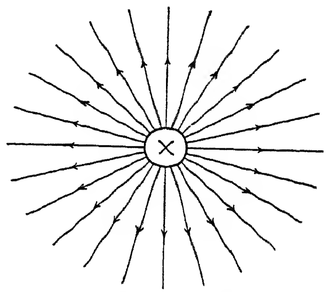


FIG. 89.—Lines of force from a single positive charge.

Difference of Potential

8. In studying mechanics, we saw that work had to be done to lift a body and that the total amount of work done in lifting a body depended only on the difference in height between the initial and final positions, not on the path through which the body moved. Conversely, if a body fell, it did work, converting its potential energy into some other form of energy. In terms of fields and field strength, we would say that in lifting a body work was done in moving it against the gravitational field and that it was moved from a point of low gravitational potential to a point of higher gravitational potential. The situation in an

electrical field is analogous. If in Fig. 89 we think of a positively charged body being brought up toward the charged sphere from a distance, it is clear that the motion is opposed at every point by a force Eq, where q is the charge on the body and E is the electric intensity at the point in question, and therefore work has to be done to overcome this force. The difference between this case and the gravitational case is that in the gravitational case our experiments are confined to a space so close to the surface

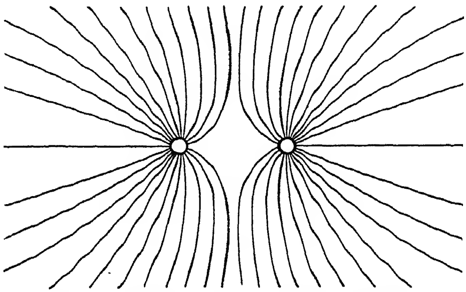


FIG. 90*a*.—Lines of force from like charges.

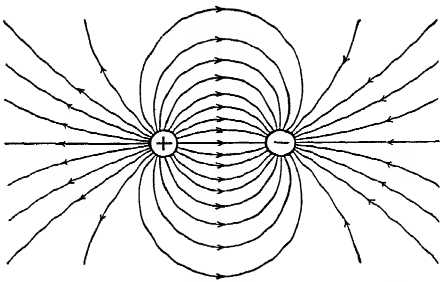


FIG. 90*b*.—Lines of force from unlike charges.

of the earth that the strength of the gravitational field does not vary appreciably, whereas in the electrical case we are interested in distances large compared with the radius of the sphere and therefore have to consider the variation of E with distance.

9. The calculation of the work done in moving a charge against such a variable force requires a simple application of integral calculus. This will be included presently for the benefit of those who have already had calculus, but first a simpler case

will be considered where the field strength is constant. This is the case of a so-called parallel-plate condenser. Such a condenser consists of two conducting planes parallel to each other and separated by a distance small compared with the dimensions of the plates. There is a positive charge on one plate and an equal negative charge on the other. The system is equivalent to that of two oppositely charged spheres shown in Fig. 90*b*, if the radii of the spheres are allowed to become infinitely large. It is

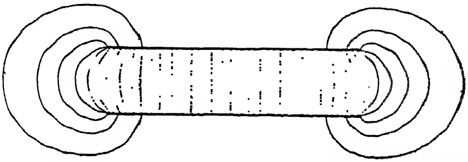


FIG. 91.—Lines of force in a parallel-plate condenser. The field in the center is uniform.

clear from that diagram that the spacing of the lines of force between the spheres will become more and more uniform as the radii of the spheres are increased and will become perfectly uniform when the spheres have grown into planes. At any rate, let us accept the result that the electric intensity in an arrangement of this sort is uniform. Now consider a small pith ball carrying a positive charge q and free to move in the space between

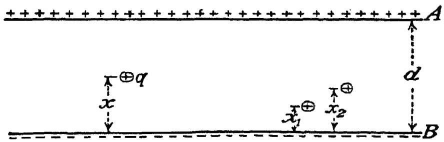


FIG. 92.—A charged body in the electric field of a parallel-plate condenser. q is the charge and x its perpendicular distance from the negative plate B. d is the separation between the positive and negative plates.

the two plates A and B. Let x be the distance of the pith ball from the negatively charged plate B and d be the separation of the plates. Let E be the electric intensity between the plates. Then there is a force Eq on the pith ball in the direction of decreasing x. (We are neglecting any gravitational force that may be present.) The analogy with gravitation is now almost perfect if the plate B is considered as the earth's surface. To increase x, work must be done, and if the pith ball is moved from x_1 to x_2, the work done is $qE(x_2 - x_1)$ just as the work done in

lifting a body from a height h_1 to a height h_2 is $mg(h_2 - h_1)$. Again if the pith ball is allowed to "fall freely" back from x_2 to x_1, it will regain the work that has been done on it, *i.e.*, the potential energy that has been put into it will be changed into kinetic energy. Again as in the gravitational case, no work is needed to move the body sideways parallel to the condenser plates since there is no force acting in this direction. Since any motion of the pith ball can be resolved into displacements parallel to the condenser plates and perpendicular to them and since only the latter use energy, the energy change for any motion of the pith ball depends only on the change in x. Thus the same amount of work is done in moving the pith ball from one plate to the other irrespective of the path by which the pith ball travels. We now introduce a notion that was not necessary in our study of gravitation but that is exceedingly useful in electricity. We call the work required to move a *unit positive charge* from the plate B to the plate A the difference of potential between B and A. In general, we define:

The difference in electrical potential between any two points is the energy per unit charge required to move a positive charge from one point to the other.

10. In terms of the units we are using, the unit of potential difference is 1 joule/coulomb of positive charge. Such a unit is called a volt and is the unit of potential difference in common everyday use. For example, suppose the plus and minus terminals of the 120-volt d.-c. laboratory supply are connected to two condenser plates some distance apart. Since we could not possibly get in practice as much as 1 coulomb of charge on a test body, let us assume that we have 1 microcoulomb of positive charge on a small insulated sphere close to the condenser plate connected to the negative side of the 120-volt line. To move it to the condenser plate connected to the positive side of the line, we shall have to do 120×10^{-6} joule of work, neglecting its weight. If the test sphere is a pith ball weighing 50 mg., this amount of work would suffice to lift it about 25 cm. in the absence of any force except gravity.

11. No definition of absolute potential has been given, only a definition of *difference of potential*. It is often convenient to refer to the difference of potential between some point in space and some standard point. It is an experimental fact that the

earth is a sufficiently good conductor to be all at the same potential. It is, therefore, what we call an *equipotential surface*, and we can arbitrarily say that it is at zero potential. For the sake of abbreviation, we can then speak of the difference of potential between any point and the earth as the potential of that point. This is a matter of convenience, and the student should realize that it is only the difference of potential that has any real meaning. Often in the mathematical study of problems in electricity, a hypothetical point at infinity is taken as the point of zero potential.

12. The student may wonder why so much emphasis is laid on these questions in the realm of electricity when they were hardly mentioned in mechanics. There are good reasons for this difference. The gravitational attractions between any masses that can be handled are so small that the gravitational field strength remains practically unaffected by anything we can do. Consequently, if no electrical forces are used, we have to control the movements of bodies by pushes and pulls transmitted by actual material substances, such as ropes, belts, and driving rods. But in the case of electricity, it is found impracticable to move electric charges around by the actual motion of the bodies carrying the charge. Save in a few instances like the Van de Graaff machine, the motions of electric charges, *i.e.*, electric currents, are controlled by electric and magnetic fields of force, that is, by "action at a distance" between different charges at rest and in motion. The whole problem of setting up currents and using them is one of setting up differences of potential at one place by doing work and using them somewhere else to furnish work.

13. In Par. 5, we pointed out that the intensity of the electric field set up by the charge on a sphere varied with the distance from the center. We want now to calculate the differences in potential between points at different distances from the sphere. Suppose we have a sphere whose center is at O and on which there is a positive charge Q. We want to calculate the difference of potential between any two points P_1 and P_2 distant r_1 and r_2, respectively, from the center of the sphere. Since the electric field is outward from O, the potential at P_1 must be greater than at P_2, *i.e.*, it will require energy to move a positive charge from P_2 to P_1; and this same amount of energy will be acquired as

kinetic energy by the same positive charge if it is allowed to move from P_1 to P_2 under the action of the electric field. The two quantities are equivalent and either one can be used as a measure of the potential difference between P_1 and P_2. The mathematical process is a little more obvious if the second process is considered. Suppose then that we start with a positive charge q at P_1. The force on it will be Eq where E is the value of the electric field at P_1. If there is no other force on it, it will move toward P_2 and in going an infinitesimal distance dr will

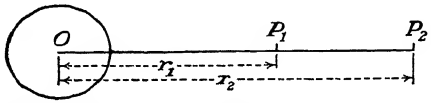


FIG. 93.—Potential as a function of the distance from a charged sphere. A fixed positive charge Q is placed at O. Work is done on a positive charge q as it moves from P_1 to P_2.

have an amount of work $Eq\,dr$ done on it. The value of E will have changed slightly so that for a second little displacement a new value of E will have to be used. This process can be continued until P_2 has been reached by a series of infinitesimal displacements each with its appropriate E and each contributing to the kinetic energy of the charge. Our problem, then, is to sum up all these contributions, and it is this process of summing up that we perform by using integral calculus. We write the sum as

$$\int_{r_1}^{r_2} Eq\,dr \tag{1}$$

which we call the definite integral of $Eq\,dr$ between the limits r_1 and r_2 (the values of r at P_1 and P_2). Now we know that E at any point distant r from O is given by $Q/4\pi\epsilon r^2$; therefore, we can rewrite our integral as

$$\int_{r_1}^{r_2} \frac{Qq\,dr}{4\pi\epsilon r^2} \tag{2}$$

where both Q and q are constant. A knowledge of calculus allows us to evaluate this integral as follows:

$$\int_{r_1}^{r_2} \frac{Qq\,dr}{4\pi\epsilon r^2} = \frac{Qq}{4\pi\epsilon}\left[-\frac{1}{r}\right]_{r_1}^{r_2} = \frac{Qq}{4\pi\epsilon}\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \tag{3}$$

Now, if q is a unit positive charge, this reduces to

$$\frac{Q}{4\pi\epsilon}\left(\frac{1}{r_1} - \frac{1}{r_2}\right)$$

which is the energy acquired by a unit positive charge allowed to move from P_1 to P_2, and this in turn is equal to the work required to move a unit positive charge from P_2 to P_1. Therefore, the difference of potential between any two points P_1 and P_2 distant r_1 and r_2 from a charged sphere is

$$V_1 - V_2 = \frac{Q}{4\pi\epsilon}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \tag{4}$$

where Q is the charge on the sphere. If one of the points, say P_2, is at infinity, we have

$$V_1 - V_\infty = \frac{Q}{4\pi\epsilon r_1}. \tag{5}$$

We cannot get more than an infinite distance away, and therefore it is natural to call the potential at infinity zero. On this basis, we would speak of the potential of any point P, at a distance r, from the center of the sphere as having a potential given by $Q/4\pi\epsilon r$ and would understand by the "potential of the point P" the difference of potential between P and a point at infinity. On this basis, the charged sphere itself has a potential of $Q/4\pi\epsilon R$ volts where R is its radius in meters.

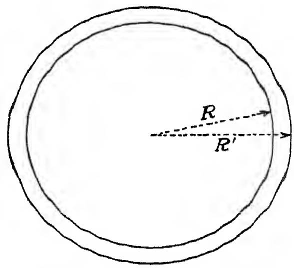


FIG. 94.—A spherical condenser.

14. In practice, we cannot isolate a charged sphere completely from all other charges and conductors. The charge on the sphere will always induce an opposite charge on the table on which it stands, on the walls of the room in which it is, etc., so that the lines of force going out from it will not all go to infinity undisturbed but will many of them terminate on induced charges on neighboring surfaces. Let us consider a particularly simple example. Suppose that a charged sphere is in the center of a spherical room whose walls are conducting and connected to earth. If the room is infinitely large, the problem reduces to

the one we have already treated. But suppose the room has a finite radius R'. Then the charge Q on the small sphere will induce a charge $-Q$ on the walls of the room. The electric intensity in the room will not be altered since we have seen that a charge on a conducting sphere produces no electric field inside that sphere; but outside the room the effect of the charge $-Q$ will just counteract the effect of the charge Q on the inner sphere so that the electric intensity will be zero (see footnote to Par. 5). In terms of lines of force, the same number of lines of force as before will start from the charge Q on the inner sphere and go out from it radially but they will all terminate on the charge $-Q$ on the walls of the enclosing room instead of extending to infinity. Now consider the potential of the inner sphere. Any charge on the earth being neglected, if a unit positive charge is brought up from infinity, it encounters no electric force until it gets inside the spherical room. Inside the room, it encounters the same electric force as before, *viz.*, $Q/4\pi\epsilon r^2$ and therefore has to have $\frac{Q}{4\pi\epsilon}\left(\frac{1}{R} - \frac{1}{R'}\right)$ joules of work done on it to get to the surface of the inner sphere. By definition, this is the difference of potential between the inner sphere and the enclosing room which is connected to the earth.* On this basis, then, we see that the effect of enclosing the original sphere by another one is to reduce its potential. Suppose we are trying to store as much positive charge on the inner sphere as possible. Each additional charge that we bring to it has to be brought against the opposing potential $\frac{Q}{4\pi\epsilon}\left(\frac{1}{R} - \frac{1}{R'}\right)$ where Q is the charge previously accumulated. Therefore, the smaller the value of $\left(\frac{1}{R} - \frac{1}{R'}\right)$, the easier it

* If the earth were uncharged, no work would be done in bringing up the unit charge from infinity to the earth, and therefore the difference of potential between the inner sphere and the earth and between the inner sphere and infinity would be the same. Actually, the earth is thought to have some resultant charge so that work is done in bringing the charge up from infinity to the earth but if we take the earth as the zero of potential this need not worry us. We can leave the question of the potential of infinity and of the difference of potential between it and an isolated charged sphere to more advanced treatises. For our purposes, it is much more to the point to discuss cases that can be realized experimentally and to take the earth as the standard of potential.

is to bring up a given amount of charge to the inner sphere. In technical language, we say that the capacity of the system is greater the smaller the value of $\left(\frac{1}{R} - \frac{1}{R'}\right)$. We shall now go on to a general discussion of condensers and capacity.

Condensers and Capacity

15. The system of two concentric conducting spheres which we have been discussing is, in effect, an arrangement for storing electric charge and electric energy. In general, any pair of conductors separated by insulating material forms such a device; a positive charge built up on one conductor induces a negative charge on the other, or vice versa; energy is required to build up the charges and this energy manifests itself in the potential difference set up between the two conductors. Any such device is called an electrical condenser, and we describe the ease or difficulty with which a given amount of charge can be put into a condenser by saying that the condenser has a large or small capacity. In more precise terms, the *capacity of a condenser is defined as the ratio between the charge on one of its two conductors and the potential difference set up in the condenser by that charge.* In mathematical language,

$$C = \frac{Q}{V} \tag{6}$$

where C is the capacity of the condenser, Q is the charge on one of its plates, and V is the potential difference set up by this charge between the two plates. In general, it is found both theoretically and experimentally that this ratio is constant for a given condenser. If the charge is doubled, the potential difference is doubled; if the charge is halved, the potential difference is halved.

16. In the system of units that we are using, a condenser is said to have unit capacity when a charge of 1 coulomb on one of its plates sets up a potential difference of 1 volt between the plates. This unit is named the *farad* after Michael Faraday. It is an enormously large capacity, illustrating the fact that a consistent logical system of units can not be "practical" in all parts of the subject. (The electrostatic unit of capacity turns out to be as ridiculously small as the farad is large.) In prac-

tice, the microfarad, *i.e.*, one-millionth of a farad, is the most commonly used unit of capacity. A condenser of 1-microfarad capacity would be raised to a potential difference of 1 volt by 1 microcoulomb of charge.

17. To get a concrete example of the size of the units of capacity, let us go back to our system of two concentric conducting spheres. According to the definitions we have just given, this system is a condenser, descriptively called a spherical condenser, and its capacity is given by the ratio of Q to V. According to the discussion of Par. 14, this is

$$C = \frac{Q}{V} = \frac{Q}{\frac{Q}{4\pi\epsilon}\left(\frac{1}{R} - \frac{1}{R'}\right)} = \frac{4\pi\epsilon RR'}{R' - R}. \tag{7}$$

Suppose that the space between the two spheres is filled with air so that $\epsilon = \epsilon_0$ and $4\pi\epsilon = 10^{-9}/9$ as we have already seen. Also let R be 1 m. and $R' - R$ be 1 cm. RR' can be taken as approximately equal to R^2 and $\frac{1}{9}$ as $\frac{1}{10}$. Making these substitutions, we get

$$\begin{aligned} C = \frac{10^{-10} \times 1}{1 \times 10^{-2}} &= 10^{-8} \text{ farad} \\ &= 10^{-2} \text{ microfarad} \end{aligned}$$

Thus we see that a spherical condenser 2 m. in diameter and with inner and outer spheres only 1 cm. apart has so small a capacity that a charge of 1 microcoulomb would set up a potential difference of 100 volts in it. If we were dealing with a single conducting sphere and calculating the capacity of the condenser formed by it and the surrounding space, we would have to set R' equal to infinity. Equation (7) would then become

$$C = 4\pi\epsilon_0 R \tag{8}$$

which shows that the capacity of a sphere of 1 m. radius is approximately 10^{-10} farad. Incidentally, this also follows directly from Eq. (5).

18. An electrical condenser is often compared with a reservoir for the storage of water. A better analogy is a gas tank. For example, think of the tank for compressed air that supplies air pressure to the rooms of a laboratory. A specific example is a cylindrical steel tank 0.50 m. in diameter and 1.40 m. high so

that its volume is about 0.275 cu. m., or 275 l. When this is filled with air at atmospheric pressure, it has 275×1.2 g. of air in it. If the pump runs for long enough to bring the pressure up to double atmospheric, the same tank will contain $275 \times 1.2 \times 2$ g. of air, and so on. The amount of air in the tank is proportional to the "capacity" (volume) of the tank and to the pressure of the air in the tank; or, conversely, the ratio of the amount of gas in the tank to the pressure is a constant characteristic of the tank and is what we call the capacity of the tank.

19. Water reservoirs and gas tanks fulfill two functions. One is the storage of material, water for the bath tub or fuel gas for the kitchen stove. The other is the storage of power, water to turn water wheels, or compressed air to run drills. The same thing is true of electrical condensers, but the uses of electric charges not in motion are few so that the storage of power is the only important function of an electrical condenser. The work that has to be done in bringing additional charge to the condenser plates after they have begun to charge up is stored in the condenser and recovered when the condenser is discharged. First, we do work to set up the difference of potential in the condenser, and then we recover this work by letting the charges on the condenser plates move away under the influence of this difference of potential. Just as the amount of work that can be obtained from a tank of compressed air depends on the volume of the tank and the pressure of the air in it, so the amount of energy that can be obtained from an electrical condenser depends on the capacity of the condenser and the potential to which it has been charged.

Capacity of a Parallel-plate Condenser.

20. The most common form of condenser consists of a series of parallel conducting plates whose length and breadth are large compared with their separation. Between the plates is some insulating material, often mica or glass, whose dielectric constant ϵ is greater than ϵ_0. In its simplest form, such a condenser consists of two parallel plates, and we shall show how the capacity of such a condenser can be calculated. We shall start from the case of the spherical condenser that we have already studied. We saw that the capacity of such a condenser was

$$C = \frac{4\pi\epsilon RR'}{R' - R} \tag{9}$$

where R and R' were the radii of the spheres and ϵ was the dielectric constant of the medium separating them. Now let R and R' be very large and very nearly equal. Then any small section cut out of the two spheres will consist of two nearly plane plates parallel to each other and a distance $R' - R$ apart. In other words, it will approximate a parallel-plate condenser. Let the area of one plate of this small section be A and the area of the surface of one whole sphere be S. Since the arrangement is perfectly symmetrical, the ratio of the capacity of the small section cut out to the capacity of the whole system will be simply A/S. Thus if we call the capacity of the small section C and the capacity of the whole system C', we have C/C' equal to A/S or $C = C'A/S$. But if R and R' are approximately equal, we can write, for either sphere, $S = 4\pi R^2 = 4\pi R'^2 = 4\pi RR'$, approximately. Therefore, substituting for $4\pi RR'$ in the numerator of Eq. (9), we have

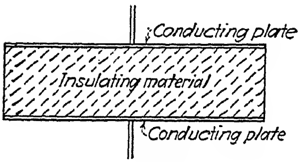


FIG. 95.—A parallel-plate condenser with capacity increased by a dielectric medium between the plates.

$$C' = \frac{\epsilon S}{R' - R}$$

and

$$C = \frac{AC'}{S} = \frac{A\epsilon}{R' - R};$$

but $R - R'$ is merely the separation of the plates of the condenser, a quantity usually called d. Substituting this symbol, we have finally for the capacity of a parallel-plate condenser

$$C = \frac{\epsilon A}{d} \tag{10}$$

where C will be given in farads if A is in square meters, d in meters, and it is evident that the dielectric constant ϵ can be expressed in farads per meter in our system of units. (*Cf.* footnote to Par. 13, Chap. X, p. 218.)

Relative Dielectric Constant

21. The general statement of Coulomb's law for the force between charges as given in Eq. (3), page 218, included the dielectric constant in the denominator. In practice, we rarely need to substitute in this equation the value of this constant for any medium other than air or vacuum. But we frequently have to compute the effect of different mediums on the capacity of condensers. For this purpose, we find it most useful to introduce the relative dielectric constant, sometimes called the specific inductive capacity. This can be defined as the ratio of the dielectric constant to the dielectric constant for vacuum or, stated mathematically, $K = \epsilon/\epsilon_0$ where K is the relative dielectric constant. Another definition, probably better because more closely related to the practical measurements, is as follows: *The relative dielectric constant of any substance is the ratio of the capacity of a given condenser when the space between the plates is filled with that substance to the capacity of the same condenser when the space between the plates is evacuated.*

22. The relative dielectric constant for a number of common substances is given in Table 12. Since these numbers are simply ratios, they will be good for any system of units. To get the value of the dielectric constant ϵ appropriate in the system of units we are using, they should be multiplied by ϵ_0, the dielectric constant for vacuum which has the value 8.84×10^{-12} farad/m. as already given.

TABLE 12.—RELATIVE DIELECTRIC CONSTANTS
(Multiply by $\epsilon_0 = 8.84 \times 10^{-12}$ farad/m. to get dielectric constants in m.k.s. practical units.)

Substance	K	Substance	K
Air	1.00059	Mica	5.6–6.0
Asphalt	2.68	Oil, insulating	2.5
Benzene	2.3	Paper	3.5
Bromine	3.18	Paraffin	2.10
Diamond	16.5	Porcelain	6
Ebonite	2.72	Shellac	3.1
Ethyl alcohol	25.8	Sulfur	4.2
Ethyl ether	4.3	Water	81.1
Glass (flint)	9.90	Wood	3
Ice (−2°C.)	2.9		

23. The effect of dielectrics in increasing the capacity of a condenser can be interpreted rather simply in terms of the electrical structure of matter. Suppose the space between the plates of a parallel-plate condenser is full of some substance, glass for example. Each molecule of the glass has both positive and negative charges in it, and the centers of gravity or average positions of the two kinds of charges are the same. Now suppose the condenser is charged, one plate positive, the other negative. This sets up an electric field which acts on the charges in the glass between the plates. The charges in the molecules are slightly displaced from their equilibrium positions in the molecules. (Since glass is a nonconductor, we know that they can not move freely.) The electrons are displaced toward the positive plate of the condenser and the positive nuclei toward the negative plate. The glass is said to be polarized. A schematic diagram is shown in Fig. 96 where each ellipse represents a molecule. In effect, the polarization produces free charges at the surface of the dielectric next to the plates, negative charges next to the positive plate, and positive charges next to the negative plate. The result is that the net charge in the neighborhood of each plate is reduced, and therefore the potential difference between the plates is reduced. For a given charge on the plates, then, the presence of a dielectric medium reduces the potential difference between the plates. Since the capacity is the ratio of the charge actually on the plates to the potential difference, the effect of the dielectric medium is to increase the capacity.

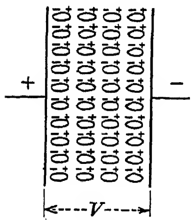


FIG. 96.—Displacement of charges in the molecules of an insulating material between the charged plates of a condenser. Each ellipse represents a molecule schematically.

Capacities in Series and in Parallel

24. In dealing with electrical circuits, we often wish to know the resultant effective capacity of a number of different condensers connected together. This could always be obtained from the general principles we have already discussed, but it is easier to use rules that we can easily deduce for the two simplest ways of connecting condensers. These two ways of connecting condensers are shown schematically in Fig. 97(*a*) and (*b*). The

condensers in Fig. 97(*a*) where each side of every condenser is connected to one or the other of two continuous conductors are said to be in parallel. It is possible to get from one of the conductors to the other by parallel routes through any one of the condensers. The condensers in Fig. 97(*b*) are said to be connected in series. The only way to get from one conductor leading into the system to the other is by going successively through the whole series of condensers.

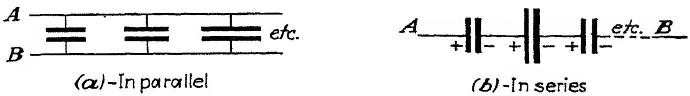


FIG. 97.—Condensers connected in parallel and in series.

25. Our problem is to get the equivalent capacity C of the whole system in terms of the separate capacities C_1, C_2, C_3, . . . , etc., of the individual condensers. Consider the condensers in parallel in Fig. 97(*a*). Let the charges on them be Q_1, Q_2, Q_3, . . . , respectively, and the potential difference be V_1, V_2, V_3, ... Since all the upper plates are connected to the conductor A, they must be at the same potential; similarly, the lower plates must be all at some one other potential. Therefore, $V_1 = V_2 = V_3 \cdots = V$, the potential difference across the whole system. The charge on the whole system is evidently the sum of the charges on the individual condensers or

$$Q = Q_1 + Q_2 + Q_3 + \cdots$$

but the total capacity $C = Q/V = (Q_1 + Q_2 + Q_3 + \cdots)/V$ but $Q_1/V = Q_1/V_1 = C_1$, etc. Therefore,

$$C = C_1 + C_2 + C_3 + \cdots \tag{11}$$

i.e., the capacity of a system of condensers connected in parallel is the sum of the capacities of the separate condensers.

26. If we turn to the case of condensers in series as shown in Fig. 97(*b*), the result is not quite so obvious. Think of the charging process. Originally there is no charge on any of the first condenser plates. Then some negative charge flows from A, leaving A with an equal positive charge. This induces an equal negative charge on the other plate of the first condenser, but this plate and the adjacent plate of the second condenser are

a connected pair of conductors insulated from the rest of the system and originally uncharged. Consequently when a negative charge is drawn to the second plate of the first condenser by induction, an equal positive charge must be left on the first plate of the second condenser. This in turn induces an equal negative charge on the opposite plate of this condenser, and so on. The result is that each condenser carries the same charge but that the only charge that flows in the whole system in the process of charging is the negative charge that left the first plate of the first condenser and entered the last plate of the last condenser. Calling this charge Q, we have

$$Q = Q_1 = Q_2 = Q_3 \cdots .$$

Now consider the potential differences. To get from A to B, every condenser must be crossed one after the other so that the work done in moving a unit positive charge from A to B must be the sum of amounts of work done in going across each condenser separately. By definition, these are the potential differences so that we have

$$V = V_1 + V_2 + V_3 + \cdots .$$

But we know that $V_1 = Q_1/C_1$, $V_2 = Q_2/C_2$, etc.; therefore,

$$\begin{aligned} V &= \frac{Q_1}{C_1} + \frac{Q_2}{C_2} + \cdots \\ &= Q\left(\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots\right). \end{aligned}$$

Also $V = Q/C$; therefore,

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \qquad (12)$$

i.e., the reciprocal of the capacity of a system of condensers connected in series is equal to the sum of the reciprocals of the capacities of the separate condensers.

Energy in a Condenser

27. We have mentioned the fact that the chief use of an electrical condenser is for storage of energy, and it is therefore interesting to inquire just how much energy such a condenser will

hold. Probably the answer can be found most easily by returning to the charged sphere discussed in Par. 14 and considering the process of charging. Evidently no work is required to bring the first little bit of charge to the sphere, but, as soon as there is any charge on the sphere, the work required to bring up additional charge is proportional to the charge already there. This steadily increases until the last little bit of charge ΔQ requires $\Delta Q \times Q(R' - R)/4\pi\epsilon RR'$ joules to bring it up or, in terms of the capacity of the spherical condenser, $\Delta Q \times Q/C$. The first little bit of charge required no work, the last little bit required $\Delta Q \times Q/C$, and the amount of work increased steadily in between; therefore, the average amount for an element of charge ΔQ was $Q\Delta Q/2C$. The total amount of work in the process of charging will be the sum of the products of each little charge brought up and the average amount of work per charge. Symbolically, total work $= \Sigma Q\Delta Q/2C = (Q/2C)\Sigma\Delta Q$ but $\Sigma\Delta Q = Q$. Therefore, the total amount of work done in the process of charging was $Q^2/2C$, and this is the amount of energy recoverable. But $Q/C = V$ so that we can write this $VQ/2$ or $V^2C/2$. This is a general result so that we can say

If a condenser of capacity C farads is charged to a potential difference of V volts, it contains $V^2C/2$ joules of electrical energy.

Electric Doublets or Dipoles

28. In Fig. 90(*b*) of this chapter, we showed the lines of electric force arising from two equal but unlike charges some distance away from each other. There is a special case of this arrangement of charges which has great importance and which can be quantitatively treated by elementary methods. It is the case where the distance separating the charges is very small compared with the distance of both charges from the points in space where their effects are to be considered. We shall find that then neither the magnitude of the charges nor their separation enters separately into the final result of our calculations; when one occurs, it is always multiplied by the other. Such an arrangement of two equal but opposite charges very close together is called an electric doublet or electric dipole. The product of the strength of one charge and the distance between the charges is called the *moment of the dipole*. A dipole can be completely represented by a vector whose magnitude is proportional to the

moment of the dipole and whose direction is the direction of the line drawn from the negative charge to the positive.

The Potential from a Dipole

29. Let us proceed to calculate the potential at any point P (*i.e.*, the difference of potential between this point and infinity) resulting from a dipole at the point O. Let the negative charge of the dipole be at A and the positive charge be at B as shown in Fig. 98. By definition, the distance AB which we shall call L is small compared with the distance r from O to P. Let the distance AP be r_1, the distance BP be r_2, and the amount of charge at A or B be Q. The potential is not a vector quantity, and the potential at any point resulting from several charges is simply the algebraic sum of the potentials calculated for each charge separately. Consequently, we can apply the results of Eq. (5) and write the potential at P as

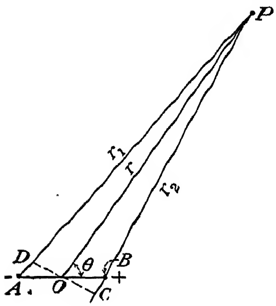


FIG. 98.—The potential from an electric dipole. r is the distance from the center of the dipole to the point P where the potential is being found. θ is the angle between r and the axis AB of the dipole.

$$V = \frac{Q}{4\pi\epsilon r_2} - \frac{Q}{4\pi\epsilon r_1}. \tag{13}$$

Draw a line DOC through O perpendicular to r. Since r is very long compared with L, this line DOC will be approximately perpendicular to r_1 and r_2 also. Let θ be the angle between r and AB. The angles DAO and OBC are then also approximately equal to θ. Therefore, r_1 which is equal to $r + AD$ can be written $r + AO \cos\theta$ or $r + \frac{1}{2}L \cos\theta$. Similarly, $r_2 = r - \frac{1}{2}L \cos\theta$. Making these substitutions in Eq. (13), we get

$$V = \frac{Q}{4\pi\epsilon}\left(\frac{1}{r - \frac{1}{2}L\cos\theta} - \frac{1}{r + \frac{1}{2}L\cos\theta}\right)$$

or

$$V = \frac{Q}{4\pi\epsilon}\left(\frac{L\cos\theta}{r^2 - \frac{L^2}{4}\cos^2\theta}\right); \tag{14}$$

but since L is very small compared to r, we can neglect $(L^2/4)\cos^2\theta$ compared with r^2 in the denominator. Therefore,

$$V = \frac{QL \cos \theta}{4\pi\epsilon r^2}. \tag{15}$$

By definition, QL is the moment of the dipole that we can call p. Substituting this quantity, we have

$$V = \frac{p \cos \theta}{4\pi\epsilon r^2} \tag{16}$$

for the electrical potential caused by a dipole of moment p at a point a distance r away in a direction making an angle θ with the line from the negative to the positive charge of the dipole.

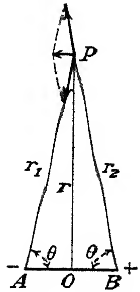


FIG. 99.—The electric field of a dipole at a point P on the perpendicular bisector of the axis AB. The components of the field perpendicular to the axis cancel so that the resultant field is parallel to the axis.

The Electric Field Set Up by a Dipole

30. By elementary methods, it is possible to calculate the electric field of a dipole only in two directions, along the axis of the dipole and perpendicular to it. These special cases follow.

1. *Perpendicular to the Axis.* From Par. 5 of this chapter, we see that the electric field at P caused by the positive charge at B is

$$\frac{Q}{4\pi\epsilon r_2^2}$$

in the direction BP and that caused by the negative charge at A is

$$\frac{Q}{4\pi\epsilon r_1^2}$$

in the direction PA. From symmetry, it is clear that the components of these two vectors along PO are equal and opposite so that the resultant field is the sum of the components parallel to AB. These will be

$$\frac{Q \cos \theta}{4\pi\epsilon r_2^2} \quad \text{and} \quad \frac{Q \cos \theta}{4\pi\epsilon r_1^2}$$

and are in the same direction. But $r_1 = r_2$ and

$$\cos \theta = \frac{AO}{AP} = \frac{L}{2r_1} = \frac{L}{2r_2};$$

therefore, the electric field E at P is given by

$$E = \frac{Q}{4\pi\epsilon}\left(\frac{L}{2r_1^3} + \frac{L}{2r_2^3}\right) = \frac{Q}{4\pi\epsilon}\frac{L}{r_1^3}. \tag{17}$$

But if P is very far away compared with the length L of the dipole, $r = r_1 = r_2$. Also we can substitute for QL the moment p of the dipole. Hence,

$$E = \frac{QL}{4\pi\epsilon r^3} = \frac{p}{4\pi\epsilon r^3}. \tag{18}$$

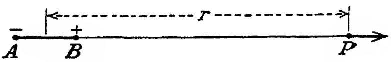


FIG. 100.—The electric field of a dipole at a point P on the axis AB of the dipole. The field at P due to each charge is along the axis so that they add algebraically to give the resultant field.

2. *Parallel to the Axis.* Evidently the field at P in Fig. 100 is in the direction BP and is given by

$$E = \frac{1}{4\pi\epsilon}\left[\frac{Q}{(r - L/2)^2} - \frac{Q}{(r + L/2)^2}\right] = \frac{Q}{4\pi\epsilon}\left[\frac{2Lr}{(r^2 - L^2/4)^2}\right],$$

but $L^2/4$ can be neglected in comparison with r if P is at a large distance. Consequently,

$$E = \frac{2LQ}{4\pi\epsilon r^3} = \frac{2p}{4\pi\epsilon r^3} \tag{19}$$

or just twice the field at an equal distance to one side of the dipole.

SUMMARY

Wherever a body is in the neighborhood of the earth, it finds itself acted on by gravity. Such a region throughout which forces are acting is called a field of force. The earth is surrounded by a gravitational field of force; a charged body is surrounded by by an electric field of force. The intensity of an electric field E is equal to the force per unit charge which would act on a positive charge at that point. The force on a charge q is Eq. The nature of an electric field corresponding to a certain distribution of electric charges is indicated by drawing "lines of force." The difference of electric potential between two points is defined as

the work per unit charge needed to move a positive charge from one point to the other. One joule of work is necessary to move 1 coulomb through a potential difference of 1 volt. The difference of potential between a point and infinity or the earth is called the potential of that point. The potential at a point P distant r from a point charge is found to be $Q/4\pi\epsilon r$.

Any system of conductors on which charges and, therefore, electric energy can be stored is called an electrical condenser. The capacity of a condenser is the ratio of the charge on it to the potential difference set up by that charge, *i.e.*, $C = Q/V$. A condenser in which a charge of 1 coulomb sets up a potential difference of 1 volt has a capacity of 1 farad. A condenser is analogous to a compressed-air tank. The capacity of a spherical condenser is $4\pi\epsilon RR'/(R' - R)$ and that of a plane condenser is $\epsilon A/d$. The relative dielectric constant of a substance is the ratio of the capacity of a condenser filled with that substance to the capacity of the same condenser filled with vacuum. Relative dielectric constants are all greater than one, a fact that is explicable in terms of the electric nature of matter. The capacity of a system of condensers in parallel is equal to the sum of their individual capacities. The reciprocal of the capacity of a system of condensers in series is equal to the sum of the reciprocals of the individual capacities. The energy stored in a condenser is $V^2C/2$.

Two equal but opposite charges very close together are called an electric dipole, and the product of one of the charges and the distance between them is called the moment of the dipole. The potential at a point P distant r from a dipole of moment is $\frac{p \cos \theta}{4\pi\epsilon r^2}$ where θ is the angle between r and the axis of the dipole. The field strengths at points distant r perpendicular to the axis and r along the axis are, respectively, $p/4\pi\epsilon r^3$ and $2p/4\pi\epsilon r^3$.

ILLUSTRATIVE PROBLEMS

1. If a charge of 4 microcoulombs is placed 12 m. from a charge of -8 microcoulombs, what will be the magnitude and direction of the electric field at each charge resulting from the other? What is the force on each charge?

According to Par. 5 the electric field resulting from a point charge q is

$$E = \frac{1}{4\pi\epsilon} \times \frac{q}{d^2}$$

where ϵ is the dielectric constant of the medium surrounding the charge and d is the distance to the point in question. E is positive if it is directed away from q (see Fig. 89). From Chap. X, Par. 13, we know that $1/4\pi\epsilon_0 = B = 9.0 \times 10^9$ m./farad or newton-m.^{2}/coulomb^2 for empty space and is approximately the same for air, which is assumed to be the medium in which the charges are unless otherwise stated.

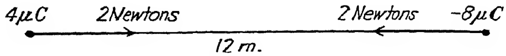


FIG. 101.—The forces on charges of 4 and −8 microcoulombs placed 12 m. apart.

Hence the electric field at the −8 microcoulomb charge caused by the 4 microcoulomb charge is

$$E = 9 \times 10^9 \frac{\text{newton-m.}^2}{\text{coulomb}^2} \times \frac{4 \times 10^{-6}\text{ coulomb}}{(12\text{ m.})^2}$$
$$= \frac{36 \times 10^3}{144} \frac{\text{newton}}{\text{coulomb}} = 250\text{ newtons/coulomb.}$$

and the positive sign shows that the field is directed away from the 4 microcoulomb charge.

The magnitude of the force on the −8 microcoulomb charge is then given by $F = Eq$ (Par. 4) or

$$F = 250 \frac{\text{newtons}}{\text{coulomb}} \times (-8) \times 10^{-6}\text{ coulomb} = -2 \times 10^{-3}\text{ newton.}$$

The minus sign indicates a force of attraction toward the other charge as shown in Fig. 101.

Similarly at the point where the 4 microcoulomb charge is located the electric field strength caused by the −8 microcoulomb charge is

$$E = 9 \times 10^9 \frac{\text{newton-m.}^2}{\text{coulomb}^2} \times \frac{(-8) \times 10^{-6}\text{ coulomb}}{(12\text{ m.})^2}$$
$$= -\frac{72 \times 10^3}{144} \frac{\text{newton}}{\text{coulomb}} = -500\text{ newtons/coulomb,}$$

and the negative sign shows that the field at the 4 microcoulomb charge is directed toward the other charge. The force on the 4 microcoulomb charge is

$$F = Eq = -500 \frac{\text{newtons}}{\text{coulomb}} \times 4 \times 10^{-6}\text{ coulomb}$$
$$= -2 \times 10^{-3}\text{ newton.}$$

The minus sign once more indicates a force of attraction. It will be noticed that the magnitudes of the two forces calculated are the same but their directions opposite, a good example of Newton's third law.

The forces on the charges could have been calculated directly from Coulomb's law but the calculation as given brings out the significance of the electric field.

2. Two charges of 2 and -4 micromicrocoulombs are placed at the extremities of the hypotenuse of an isosceles right triangle. The length of each of the equal sides is 20 cm. Find the electric field strength at the right-angle vertex of the triangle.

The electric field strength at the vertex of the triangle is the resultant of those produced by the two charges separately. These must be found first and then combined by vector addition. Let E_1 be the field caused by the 2×10^{-12} coulomb charge and E_2 be that caused by the -4×10^{-12} coulomb charge.

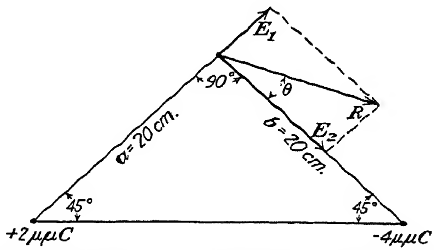


Fig. 102.—The electric field of two charges at a point on the perpendicular bisector of the line joining the two charges. $\mu\mu C$ stands for micromicrocoulombs.

Then by a calculation like that in Prob. 1 we find

$$E_1 = 9 \times 10^9 \frac{\text{newton-m.}^2}{\text{coulomb}^2} \times \frac{2 \times 10^{-12} \text{ coulomb}}{(0.2 \text{ m.})^2} = \frac{1.8 \times 10^{-2}}{4 \times 10^{-2}} \frac{\text{newton}}{\text{coulomb}}$$
$$= 0.45 \text{ newton/coulomb.}$$

This component of field strength is directed outward as shown in Fig. 102.

Similarly the -4×10^{-12} coulomb charge gives a field

$$E_2 = 9 \times 10^9 \frac{\text{newton-m.}^2}{\text{coulomb}^2} \times \frac{(-4) \times 10^{-12} \text{ coulomb}}{(0.2 \text{ m.})^2}$$
$$= -0.90 \text{ newton/coulomb.}$$

The negative sign shows that this component of field strength is directed toward the charge as shown in Fig. 102.

Finally, we must compute the magnitude and direction of the resultant field strength R by vectorial addition of E_1 and E_2. Since E_1 and E_2 are at right angles, the magnitude of their resultant is given immediately by

$$R = [E_1^2 + E_2^2]^{\frac{1}{2}} = \left[\left(0.45 \frac{\text{newton}}{\text{coulomb}}\right)^2 + \left(0.90 \frac{\text{newton}}{\text{coulomb}}\right)^2\right]^{\frac{1}{2}}$$
$$= 1.006 \text{ newtons/coulomb.}$$

The angle θ which R makes with the side of the triangle is given by

$$\tan \theta = \frac{E_1}{E_2} = \frac{0.45 \text{ newton/coulomb}}{0.90 \text{ newton/coulomb}} = 0.5$$

from which

$$\theta = 27°.$$

Thus, at the vertex of the triangle the electric field strength, or force per unit charge which would act on a positive charge at that point, is 1.006 newtons/coulomb in a direction making an angle of 27° with the side of the triangle as shown in Fig. 102.

3. If a 10-g. ball carrying a charge of 5 microcoulombs is placed in an electric field of 20 newtons/coulomb, what will be its initial acceleration due to the field?

The force acting on the ball is the product of the electric field strength and the charge on the ball.

$$\begin{aligned} F &= 20 \frac{\text{newtons}}{\text{coulomb}} \times 5 \times 10^{-6} \text{ coulomb} \\ &= 10^{-4} \text{ newton.} \end{aligned}$$

This force will give to a mass of 10 g. an acceleration of

$$a = \frac{F}{m.} = \frac{10^{-4} \text{ newton}}{10 \times 10^{-3} \text{ kg.}} = 10^{-2} \text{ m./sec}^2. = 1 \text{ cm./sec}^2.$$

4. Two condensers of capacities 3 and 5 microfarads are each charged to a potential of 100 volts. They are then connected in parallel, the positive plate of one being connected to the negative plate of the other and vice versa. What are the charge and difference of potential of each condenser?

The initial charge on each condenser is

$$\begin{aligned} Q_1 &= V_1C_1 = 100 \text{ volts} \times 3 \times 10^{-6} \text{ farad} \\ &= 300 \text{ microcoulombs} \\ Q_2 &= V_2C_2 = 100 \text{ volts} \times 5 \times 10^{-6} \text{ farad} \\ &= 500 \text{ microcoulombs.} \end{aligned}$$

The charges partially neutralize each other when the condensers are connected in parallel, so that the total charge on both condensers is 500 microcoulombs − 300 microcoulombs = 200 microcoulombs. The capacity of both condensers in parallel is

$$C = C_1 + C_2 = 3 \; \mu f + 5 \; \mu f = 8 \mu f.$$

The potential difference between the plates of the pair considered as one large condenser is, therefore,

$$V = \frac{Q}{C} = \frac{200 \times 10^{-6} \text{ coulomb}}{8 \times 10^{-6} \text{ farad}} = 25 \text{ volts.}$$

The charge on the 3-μf condenser is

$$Q_1 = VC_1 = 25 \text{ volts} \times 3 \times 10^{-6} \text{ farad} = 75 \times 10^{-6} \text{ coulomb}.$$

The charge on the 5-μf condenser is

$$Q_2 = VC_2 = 25 \text{ volts} \times 5 \times 10^{-6} \text{ farad} = 125 \times 10^{-6} \text{ coulomb}.$$

5. Find the capacity of a condenser made of two plates each of area 2 m². with a dielectric of glass 2 mm. thick between them.

If we take 8 as the specific inductive capacity of glass, the capacity of the condenser is given by Eq. (10), page 240,

$$C = \frac{\epsilon A}{d} = \frac{\dfrac{1}{4\pi \times 9 \times 10^9} \dfrac{\text{farads}}{\text{m.}} 8 \times 2 \text{ m}^2.}{2 \times 10^{-3} \text{ m.}}$$

$$= \frac{2}{9\pi} \times 10^{-6} \text{ farad} = 7.07 \times 10^{-8} \text{ farad} = 0.0707 \ \mu f.$$

PROBLEMS

1. Calculate the gravitational field at a point 1,000 mi. above sea level directly over the tip of the Washington monument, directly over the Empire State building. This is known as a radial field of force since it is independent of the direction and depends only on the distance from the center of the earth.

2. Compute the electric field strength at a distance of 1 m. from unit charge.

3. Three equal and like charges are at the corners of an equilateral triangle each side of which has a length L. What is the field strength (*a*) in the middle of the triangle? (*b*) in the middle of one side? (*c*) at a perpendicular distance L away from the middle of one side?

4. Equal and opposite charges $+q$ and $-q$ are located on small spheres whose centers are l m. apart. What are the direction and magnitude of the electric field L m. from the midpoint between the spheres, (*a*) along the line joining the spheres? (*b*) perpendicular to the line joining the spheres? Let L be larger than l. What acceleration would this electric field in (*a*) cause in a body of mass m kg., having a charge Q coulombs?

5. How much charge would there have to be on the earth to make a pith ball of mass 10 mg. and charge 10 microcoulombs float in a vacuum? If this charge were distributed uniformly over the earth's surface, what would be the charge per unit area?

6. A mass of 5 kg. is placed in a constant gravitational field of 15 newtons/kg. What will be the acceleration of the mass due to the field?

7. A 50-mg. pith ball carrying a charge of 20 microcoulombs is placed in a constant electric field of strength 100 newtons/coulomb. Find the acceleration due to the field.

8. A uniform electric field in a parallel-plate condenser has an intensity of 8 newtons/coulomb. What force would act on a charge of 50 microcoulombs placed in this field?

9. Two joules of work are done in carrying a charge of 10 microcoulombs from infinity to a certain point. What is the potential of that point?

10. What are the values of the possible capacities that can be made from three condensers with capacities of 1, 2, and 3 microfarads.

11. Four condensers of capacities $C_1 = 12$, $C_2 = 3$, $C_3 = 24$, $C_4 = 5$ microfarads are connected across a 3,000-volt battery as in Fig. 103. (*a*) Find the voltage across each condenser. (*b*) Find the charge on each condenser. (*c*) If the charged system of condensers is now removed from the battery and the connections altered to put all four condensers in series, what will be the total potential difference across the bank?

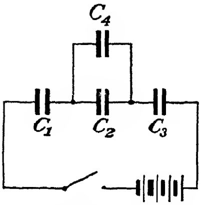


FIG. 103.—A bank of condensers connected to a 3000-volt battery. Prob. 11.

12. Compute the capacity of a sphere 2 m. in diameter.

13. A conducting sphere of radius 10 cm. carries a charge of 0.1 microcoulomb. It is temporarily connected by a long wire to an uncharged conducting sphere of radius 5 cm. and the spheres are then separated. How much charge is on each?

14. The radius of the Van de Graaff spheres in the original machine is 0.30 m. The belts move at the rate of approximately 0.30 m./min. If the sphere can be raised to a potential of 100,000 volts in 3 min., how much charge does the belt carry per unit length? What is the approximate power that the motor must supply in the last fraction of a second of the charging operation? (Assume that only the upgoing belt carries charge, the downcoming one being neutral.)

15. Suppose one of the Van de Graaff spheres charged to +100,000 volts above ground and standing isolated. Suppose a pith ball of mass $\frac{1}{2}$ g. 1 m. below the surface of the sphere has a charge of q coulombs. How great must q be for the pith ball to be attracted to the sphere, and with what velocity will it strike? (Neglect air resistance.)

16. How much charge must be added to an uncharged condenser having a capacity of 10 microfarads to cause a difference of potential of 500 volts between the plates?

17. In a parallel-plate condenser, each plate has an area of 2 sq. m. If they are separated by a 5-mm. piece of glass which is then removed, what is the change in the capacity of the condenser?

18. A parallel-plate condenser has two plates of area 1 sq. m. with a plate of glass 12 mm. thick between them. If the condenser has a charge of 10 microcoulombs, find the change in energy stored in the condenser and also the change in potential of the condenser when the glass plate is removed.

19. What is the electric field strength at a point 2 m. from an electric dipole of moment 5 coulomb-cm. if the line from the point to the center of the dipole is perpendicular to the axis of the dipole.

20. Two electric dipoles each have a moment of 2×10^{-6} coulomb-m. One dipole is placed on the axis of the other with their centers 20 cm. apart. What angle must the second dipole make with respect to the axis of the first in order that it experience a torque of 2.0 newton-meters?

21. An electric dipole of moment 2×10^{-8} coulomb-m. lies with its axis along the perpendicular bisector of the axis of another dipole of moment 3×10^{-6} coulomb-m. If their centers are 30 cm. apart, compute the torque on each due to the other.

22. An electric dipole lies on the perpendicular bisector of the axis of an equal dipole with its axis at an angle of 30° with the bisector. Their centers are 40 cm. apart. What must be the moments of the dipoles in order that the one on the bisector should experience a torque of 10^{-3} newton-m.?

CHAPTER XII

THE ELECTRONIC CHARGE

1. In previous chapters, various experiments have been interpreted in terms of the motion of electrons whose existence has been assumed. Though we are not yet in a position to describe all the experiments that are needed to determine the nature of these electrons, we can give some of the results that prove their existence and even determine the amount of charge which each of them has. First, let us consider the problem in general. The electrical experiments that have been described so far have all been known for a long time and have been interpreted in various ways. One theory which is quite satisfactory for their interpretation is that electricity is a weightless fluid that can move around freely in conductors and less freely in insulators. Whether there are two fluids, one positive and one negative, or whether the two kinds of charge result from an excess or a deficit of the same fluid need not concern us. Once the atomic theory of matter is firmly established so that we know that a fluid, like water, which appears to be perfectly homogeneous and continuous, is really made up of large numbers of discrete particles, the question naturally arises whether electric charge is really continuous or is made up of large numbers of small particles of charge. If electricity is atomic in nature, are all the atoms of electricity the same or are there several kinds or perhaps even an infinite number of kinds? Are these electrical particles weightless, and if not how much do they weigh? These are some of the questions that we shall try to answer, and as in the treatment of previous topics, we shall concern ourselves more with simplicity in our answers than with history or rigorous logic.

2. In our introductory chapters on the atomic theory of matter, we presented several different types of evidence for the existence of atoms and for their state of motion. There are the results of the study of chemical reaction, vast quantities of data requiring careful analysis which lead to the atomic hypothesis.

There are effects like the Brownian motion and the determination of molecular velocities which are direct evidence of the existence and motion of atoms. And, finally, there is the success of the kinetic theory which assumes the existence and motion of atoms. There is a rough parallel in the evidence for the existence of electrons. A study of the effects of the passage of electric current through solutions, called electrolysis, is one source of information. Another is the motion of the charged particles in an electrical discharge in gases at low pressures or the effects of changing the charge on small droplets as in Millikan's experiment. Finally, the most convincing evidence is the success of the whole theory of the electrical structure of matter which involves the assumption of small units of charge.

3. Historically and logically, the results of electrolysis should be given first and this would be consistent with our development of the atomic theory. But electrolysis involves the idea of electric current, and we prefer to introduce the idea of electric current in terms of the motion of electrons. Consequently, we want to demonstrate the existence of electrons without using electric currents. This can be done easily by discussing Millikan's famous experiment on the determination of the charge of the electron. But before beginning the account of this experiment, we had better point out clearly what is its object. Probably the student has learned in his school physics that electrons are small very light particles carrying a certain negative charge and that positive charges are carried by protons, or positive ions, that are much heavier. This is correct, and we shall have much to say along these lines in later chapters. In this chapter, we are concerned not with the carriers of the positive and negative charges but with the nature of the charges themselves, with whether they are made up of "atoms" of charge, and if so what are the smallest charges that can be obtained. We shall not attempt to draw any conclusions about the mass of the smallest particles which may carry the charges. Also following Millikan's usage, we shall occasionally use the word "electron" for an "atom" of charge whether it is positive or negative and not merely for the smallest known carrier of this amount of negative charge.

4. The principle of this experiment can be understood in terms of the description of a crude model (Fig. 104) suitable for demon-

stration to a large class. Take two sheets of tin about two meters square, and mount them horizontally about a meter apart on insulating supports. Connect them to the terminals of a Van de Graaff machine. Inflate a toy balloon, and coat it with some conducting paint until its average density is nearly the same as air. Then put it between the tin sheets. It will bounce back and forth between them. These two tin sheets form the plates of a large parallel-plate condenser so that when the Van de Graaff machine is running they are oppositely charged and attract or repel any charged body in between them. In terms of electric field and potential difference, the Van de Graaff machine produces a potential difference V between the plates which sets up

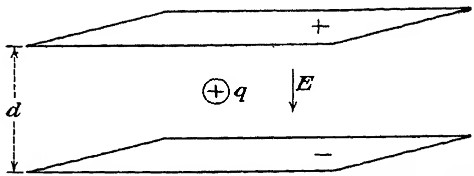


FIG. 104.—The motion of a charged balloon between two charged planes.

an electric field $E = V/d$ in the region between them, if d is the distance separating the plates. As in Par. 4, page 227, this field acts with a force Eq on any body carrying a charge q in the region between the plates. Suppose a conducting balloon carrying a positive charge $+q$ gets between the plates and that the upper plate is positive. The balloon is pulled down to the negative lower plate. (If it has no initial charge, it will rise or fall until it strikes one plate and becomes charged.) If it strikes the plate, it loses its positive charge and acquires a negative one so that it is repelled toward the upper plate. Striking this, it reverses its charge again and is driven downward. In this way, the balloon can be made to bounce back and forth between the plates. But it is found that by regulating the voltage (*i.e.*, partly discharging the upper plate to ground), it is possible to hold the balloon in equilibrium between the plates. That is, the electric forces can be made just to balance the gravitational forces. What are the gravitational forces? There is of course the weight of the balloon

$$mg = V\sigma g = \tfrac{4}{3}\pi r^3\sigma g$$

where σ is the average density of the balloon, V its volume, and r its radius. If this were the only force on a balloon, it would never rise off the ground. There is also the buoyant force of the displaced ai·, and by Archimedes' principle this is

$$V\rho g = \tfrac{4}{3}\pi r^3 \rho g$$

where ρ is the density of air. (Archimedes' principle should be familiar.) The resultant gravitational force is, therefore,

$$Vg(\sigma - \rho) = \tfrac{4}{3}\pi r^3(\sigma - \rho)g$$

and if this is positive, the balloon will fall, if negative it will rise. The total combined effect of the gravitational and electric forces is, then,

$$\tfrac{4}{3}\pi r^3(\sigma - \rho)g + Eq \qquad (1)$$

and if this is just equal to zero, the balloon does not move. If all the quantities in this expression are measured except the charge q, it can be determined. But it is difficult to keep the balloon exactly in equilibrium. Therefore, it may be necessary to observe how the balloon moves when the forces do not add up exactly to zero and to get from these observations the charge on the balloon. Unfortunately, this is not so simple since the frictional resistance of the air to the motion of the balloon through it is comparable to the other forces acting on it. This same effect will be found to play an important part in Millikan's experiment so that we had better discuss it fully before going further.

Terminal Velocity

5. As we saw in Chap. VI, the speed with which a parachute jumper hits the earth does not depend on the height from which he jumps so long as there has been ample opportunity for his parachute to open and to slow him down. The gravitational force tends to speed him up, whereas the resistance of the air tends to slow him down. But the gravitational force is a constant quantity, whereas the air resistance increases as his speed increases. Consequently, his speed increases until the air resistance is just equal to his weight. The resultant force on him is then zero so that his acceleration also becomes zero. This result is typical of a general one, according to which the viscous or frictional resistance of any fluid to the motion of any

body through it is proportional to the speed with which the body is moving. Therefore, any body moving through any fluid under the action of any constant force reaches a constant speed characteristic of the mass and shape of the body, the viscosity of the fluid, and the size of the force acting. The balloon about which we have been talking furnishes a good example of this effect. It is rather hard to study its motion in so small a space as the region between the condenser plates, but we

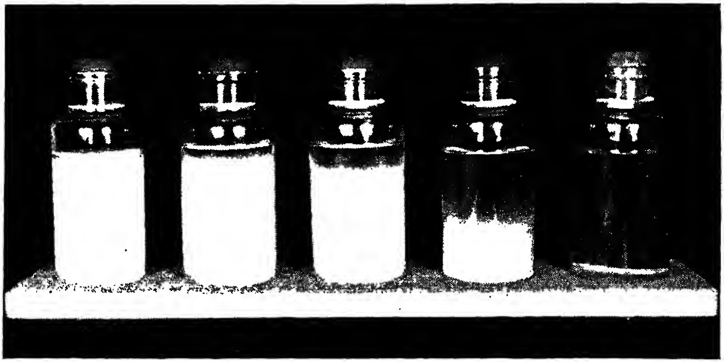


FIG. 105.—Settling of quartz particles in water. The bottles were all shaken simultaneously a few minutes before this picture was taken. The size of the particles diminishes from a diameter of between 80 and 100 millionths of a meter for the particles in the right-hand bottle to between 10 and 15 millionths of a meter for those in the left.

can eliminate the electrical effects and simply watch it rise or fall between the floor and the ceiling under the action of its weight and the buoyancy of the displaced air. Even with the disturbing effects of inevitable air currents and without any attempt at accurate timing, it will be obvious that it does not move with acceleration; it is rising no faster at the top of the room than at the bottom. In other words, it soon reaches a constant terminal velocity.

Stokes' Law

6. In certain particular cases, it is possible to predict what the terminal velocity of a body will be. This was done theoretically by Stokes for a small sphere moving in a continuous viscous medium. He found that the resistance offered by such a uni-

form fluid to the motion of a sphere was directly proportional to the first power of the velocity and was in fact given exactly by the relation

$$R = 6\pi\eta av \tag{2}$$

where R is the resistance offered by the medium whose coefficient of viscosity is η to the motion of the sphere of radius a moving with a constant velocity v. If this relation is applied to a sphere of density σ, moving under the action of gravity through a medium of density ρ, we must have the resistance just equal to the weight minus the buoyant force or

$$R = 6\pi\eta av = V(\sigma - \rho)g$$

where V is the volume of the sphere. If we express V in terms of a and substitute, we get

$$6\pi\eta av = \tfrac{4}{3}\pi a^3(\sigma - \rho)g$$

which we can solve for v, getting

$$v = \frac{2}{9}\frac{a^2g}{\eta}(\sigma - \rho). \tag{3}$$

From Stokes' law expressed in this form, we can calculate the radius of a droplet of liquid falling through a gas if we observe its velocity and know the viscosity of the gas, the density of the drop, and density of the gas.

Twofold Object of Millikan's Experiment. His Apparatus

7. The discussion of Millikan's experiment falls naturally into two parts, first the proof that all electric charges are built up out of small atoms of charge which we may as well call electrons from the start, and second the exact measurement of the magnitude of the charge of an electron. It is to be understood that Millikan by no means originated the idea of the existence of the electron. Not only had the idea been suggested, but much work had been done which showed that it was probably correct and that the charge on the electron was very small indeed, of the order of 10^{-19} coulomb. His problem therefore was to work with such minute charges that he might be able to detect changes in them which would indicate a gain or loss of several or even one

electron, just as in the Brownian movement we work with such minute masses that we can observe the changes in motion caused by the impact of a few molecules.

8. In Millikan's experiment, the balloon of our crude model was replaced by a tiny drop of oil, and the condenser plates were greatly reduced in size and improved in quality. Also the

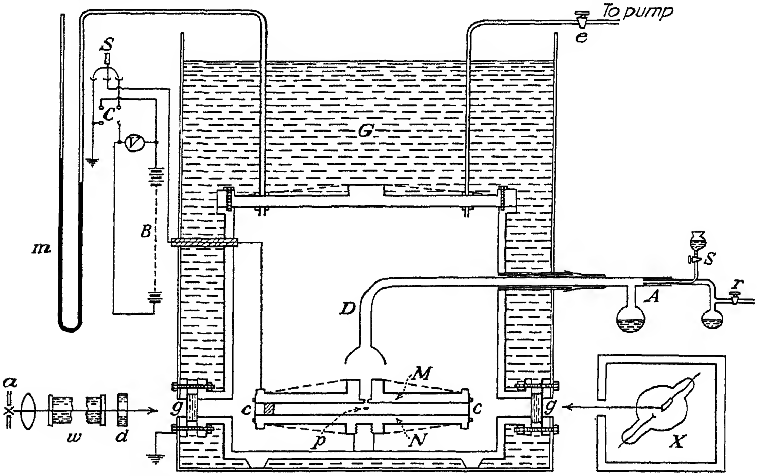


FIG. 106.—Millikan's apparatus for the determination of the charge on the electron. (*After Millikan's drawing.*)

voltage was supplied by a set of storage batteries instead of a Van de Graaff machine. The oil droplet was observed as it moved up and down under the action of gravity and the electric field. By arranging to have plenty of ions (charged molecules) in the space between the condenser plates, he made it probable that the droplet would gain or lose charge while it was under observation. This caused a change in the force on it and a consequent change in its velocity. By observing these changes, he was able to make deductions about the nature of the charge gained and lost. The details of his apparatus are best given in his own words.

9. In order to compare the charges on different ions, the procedure adopted was to blow with an ordinary commercial atomizer an oil spray into the chamber D. The air with which this spray was blown was first to be rendered dust-free by passage through a tube containing glass

wool. The minute droplets of oil constituting the spray, most of them having a radius of the order of one-thousandth of a millimeter, slowly fell in the chamber D and occasionally one of them would find its way through the minute pinhole p in the middle of the circular brass plate M, 22 cm. in diameter, which formed one of the plates of the air condenser. The other plate N was held 16 mm. beneath it by three ebonite posts. By means of the switch S these plates could be charged, the one positively and the other negatively, by making them the terminals of a 10,000-volt storage battery B, while throwing the switch the other way (to the left) short-circuited them and reduced the field between them to zero. The oil droplets which entered at p were illuminated by a powerful beam of light which passed through diametrically opposite windows in the encircling ebonite strip c. As viewed through a third window in c on the side toward the reader, it appeared as a bright star on a black background. These droplets which entered p were found in general to have been strongly charged by the frictional process involved in blowing the spray, so that when the field was thrown on in the proper direction they would be pulled up toward M. Just before the drop under observation could strike M, the plates would be short-circuited and the drop allowed to fall under gravity until it was close to N, when the direction of motion would be again reversed by throwing on the field. In this way, the drop would be kept traveling back and forth between the plates. The first time the experiment was tried, an ion was caught within a few minutes, and the fact of its capture was signaled to the observer by the change in the speed with which it moved up when the field was on. The significance of the experiment can best be appreciated by examination of the complete record of one of the early experiments when the timing was done merely with a stop watch.

10. The column headed t_g in the following table gives the successive times which the droplet required to fall between two fixed cross hairs in the observing telescope whose distance apart corresponded in this case to an actual distance of fall of 0.5222 cm. It will be seen that these numbers are all the same within the limits of error of a stop-watch measurement. The column marked t_F gives the successive times which the droplet required to rise under the influence of the electrical field produced by applying, in this case, 5,051 volts of potential difference to the plates M and N. It will be seen that, after the second trip up, the time changed from 12.4 to 21.8, indicating, since in this case the drop was positive, that a negative ion had been caught from the air. The next time recorded under t_F, namely, 34.8, indicates that another negative ion had been caught. The next time, 84.5, indicates the capture of still another ion. This charge was held for two trips, when the speed changed back again to 34.6, showing that a positive ion had now been caught which carried

precisely the same charge as the negative ion which before caused the inverse change in time, *i.e.*, that from 34.8 to 84.5.*

TABLE A

t_g	t_F	t_g	t_F
13.6	12.5	13.7	34.6
13.8	12.4	13.8	21.9
13.4	21.8	13.6	
13.4	34.8	13.5	
13.6	84.5	13.4	
13.6	85.5	13.8	
13.7	34.6	13.4	
13.5	34.8		
13.5	16.0	Mean 13.595	
13.8	34.8		

Millikan's Experiment. The Atomicity of Electricity

11. Before quoting further from Millikan, it may be well to insert some explanatory remarks. It is clear that the oil drop in this experiment plays the role of the balloon in our crude model. It moves with a terminal velocity which will be proportional to the force acting on it as long as the size and shape of the drop do not change. But there is an interesting difference which we might note in passing. With the balloon, we made the terminal velocity under the action of gravity very small by reducing the average density of the balloon until it was almost equal to that of air, so that the factor $(\sigma - \rho)$ in Eq. (3) was very small. But the density of oil is much greater than that of air. Why then does not the drop fall rapidly? It is because the a^2 in Eq. (3) is so small. In other words, the fancy expression that we have called Stokes' law is merely an exact statement of the fact that big drops fall more rapidly than little ones, a principle illustrated every time the fog droplets of a cloud condense into raindrops. Actually, the density of the oil is so much larger than that of air that the buoyant effect of the displaced air can be entirely neglected in this part of Millikan's experiment. Consequently, the only forces acting on the drop besides air friction are *mg*

* Excerpts from Robert A. Millikan, "Electrons + and −, Protons, Photons, Neutrons, and Cosmic Rays," 1935. Reprinted by permission of the University of Chicago Press.

downwards and Fe_n upward, where now we follow Millikan's notation using F for the electric field strength and e_n for the charge on the drop. If we call the downward velocity of the drop under gravity alone v_1 and the upward velocity under the resultant action of gravity and the electric field v_2, then we have from Stokes' law that each is proportional to the force acting and proportional by the same factor. Therefore, the ratio of the velocities is equal to the ratio of the forces, *i.e.*,

$$\frac{v_1}{v_2} = \frac{mg}{Fe_n - mg}. \tag{4}$$

Solving for e_n, we have

$$e_n = \frac{mg}{Fv_1}(v_1 + v_2). \tag{5}$$

The negative sign is used for mg in the denominator of (4) because, for convenience, v_2 will be taken as positive when the drop is going up, in the direction of F, and v_1 will be taken as positive when it is going down in the direction of g.

12. Consider a drop like the one used to get the data in Table A. The value of e_n evidently changes in the course of the experiment as indicated by the different values of t_F. This we have explained as caused by the capture or loss of an ion by the droplet. Assuming that the mass of the ion is negligible compared with the mass of the drop but that the charge is not, we have

$$e'_n = \frac{mg}{Fv_1}(v_1 + v'_2) \tag{6}$$

where e'_n is the new charge on the drop and v'_2 is the corresponding new velocity. But the difference between e_n and e'_n is equal to the charge on the ion which we may call e_i. We have therefore

$$e_i = e'_n - e_n = \frac{mg}{Fv_1}(v'_2 - v_2). \tag{7}$$

For a given drop, mg/Fv_1 is a constant which need not be evaluated for our present purpose. It is sufficient that the preceding equation shows the charge on the ion to be proportional to the change in the upward velocity. If, therefore, we can show that the latter change always has the same value or an integral

multiple of that value, we have shown that the charge on every ion has the same value or an integral multiple of it. The values of v_2 are found by dividing the distance through which the drop moves, 0.522 cm., by the times t_F in Table A; the changes in v_2 by subtracting successive values. Resuming the quotation from Millikan,

The successive values of v_2 and of $(v_2' - v_2)$ are shown in Table B. Values of t_F used here are means from Table A.

TABLE B

v_2	$(v_2' - v_2) \propto e_i$
0.5222/12.45 = 0.04194	
0.5222/21.85 = 0.02390	0.01804 ÷ 2 = 0.00902
0.5222/34.7 = 0.01505	0.00885 ÷ 1 = 0.00885
0.5222/85.0 = 0.006144	0.00891 ÷ 1 = 0.00891
0.5222/34.7 = 0.01505	0.00891 ÷ 1 = 0.00891
0.5222/16.0 = 0.03264	0.01759 ÷ 2 = 0.00880
0.5222/34.7 = 0.01505	0.01759 ÷ 2 = 0.00880
0.5222/21.85 = 0.02390	0.00885 ÷ 1 = 0.00885

13. It will be seen from the last column that, within the limits of error of a stop-watch measurement, all the charges captured have exactly the same value save in three cases. In all of these three, the captured charges were just twice as large as those appearing in the other charges. Relationships of exactly this sort have been found to hold absolutely without exception, no matter in what gas the drops have been suspended or what sort of droplets were used upon which to catch the ions. In many cases, a given drop has been held under observation for 5 or 6 hr. at a time and has been seen to catch not 8 or 10 ions, as in the above experiment, but hundreds of them. Indeed, I have observed, all told, the capture of many thousands of ions in this way; and in no case have I ever found one the charge of which, when tested as above, did not have either exactly the value of the smallest charge ever captured or else a very small multiple of that value. *Here, then, is direct, unimpeachable proof that the electrical charges found on ions all have either exactly the same value or else small exact multiples of that value.*

14. The foregoing results show that the charge gained or lost by the drop during its sojourn between the plates is always an integral multiple of the same charge. But what about the charge on the drop in the first place? The bearing of the data on this question can be interpreted as follows:

Dividing Eq. (5) by (7), we get

$$\frac{e_n}{e_i} = \frac{(v_1 + v_2)}{(v'_2 - v_2)}. \tag{8}$$

Now, if the smallest observed value for $(v'_2 - v_2)$ corresponds to an ion with a single elementary charge, the ratio in (8) when that value of $(v'_2 - v_2)$ is put in should be a whole number since the charge on the drop must be a whole number of elementary charges. Reversing the argument, if the ratio of Eq. (8) always comes out to be a whole number, it supports the belief in the existence of elementary charges of uniform size.

15. From the data of Table A, we have

$$v_1 = 0.5222/13.59 = 0.03842,$$

and the first value of v_2 from Table B is 0.04194. Adding these two values, we get $v_1 + v_2 = 0.08036$. Dividing this by the average of the last column of Table B, the smallest value of $v'_2 - v_2$, we get 9.05 which is a whole number within the limits of error of the observations. The repetition of this procedure with many other drops and in more accurate experiments shows that such a whole number is not accidental.

16. Millikan presents data of the same sort in a slightly different way. If the ratio in Eq. (8) is a whole number it can be written

$$\frac{e_n}{e_i} = \frac{v_1 + v_2}{v'_2 - v_2} = n$$

or

$$v_1 + v_2 = n(v'_2 - v_2).$$

We can therefore make a table of integral multiples of the minimum value of $(v'_2 - v_2)$ determined experimentally. Then we can compare the values of $v_1 + v_2$ observed for the same drop with the numbers in the table. For such a comparison, it is not necessary to use the velocities themselves in centimeters per second but merely numbers proportional to them. This is what Millikan has done in Table C where the second column contains numbers proportional to $n(v'_2 - v_2)$ and the third, numbers proportional (by the same factor) to $v_1 + v_2$. In quoting Millikan again,

It may be of interest to introduce one further table (Table C) arranged in a slightly different way to show how infallibly the

atomic structure of electricity follows from experiments like those under consideration.

TABLE C

n	$4.917 \times n$	Observed charge	n	$4.917 \times n$	Observed charge
1	4.917	.....	10	49.17	49.41
2	9.834	.....	11	54.09	53.91
3	14.75	.....	12	59.00	59.12
4	19.66	19.66	13	63.92	63.68
5	24.59	24.60	14	68.84	68.65
6	29.50	29.62	15	73.75	
7	34.42	34.47	16	78.67	78.34
8	39.34	39.38	17	83.59	83.22
9	44.25	44.42	18	88.51	

17. In this table 4.917 is merely a number obtained precisely as above from the change in speed due to the capture of ions and one which is proportional in this experiment to the ionic charge. The column headed $4.917 \times n$ contains simply the whole series of exact multiples of this number from 1 to 18. The column headed "Observed charge" gives the successive observed values of $(v_1 + v_2)$. It will be seen that during the time of observation, about four hours, this drop carried all possible multiples of the elementary charge from 4 to 18, save only 15. *No more exact or more consistent multiple relationship is found in the data which chemists have amassed on the combining powers of the elements and on which the atomic theory of matter rests than is found in the foregoing numbers.*

18. Such tables as these—and scores of them could be given—place beyond all question the view that an electrical charge wherever it is found, whether on an insulator or a conductor, whether in electrolytes or in metals, has a definite granular structure, that it consists of an exact number of specks of electricity (electrons) all exactly alike, which in static phenomena are scattered over the surface of the charged body and in current phenomena are drifting along the conductor.

19. Millikan then goes on to discuss a number of interesting by-products of his experiment, one of which showed that positive and negative charges were built of units of equal size. That is, the smallest positive charge that he ever observed was just equal in magnitude to the smallest negative charge he ever observed. Whether the smallest particle with which such charges were ever

associated had the same mass in the case of positive and negative charges is another question.

Millikan's Experiment. The Exact Determination of the Electronic Charge

20. It is not desirable in an elementary book to go into the details of the experimental precautions and theoretical corrections that were necessary before Millikan arrived at his final value of the charge on the electron. But it may be of interest to see how we can arrive at a rough value from the data that have already been given. Suppose we go back to Eq. (7). In using it before, we considered the factor mg/Fv_1 merely as a factor of proportion; but if we knew its value, we could immediately calculate the absolute value of e_i. There is no difficulty about measuring F which is simply the potential difference between the plates divided by their separation. g we know, and v_1 we have measured; but what about m? It sounds fairly easy but it is not. These drops are much too small to weigh with a chemical balance. We have to resort to Stokes' law. In fact, we introduced the discussion of Stokes' law for this very purpose. It told us [Eq. (3)] that the velocity of free fall of a drop like this was given by

$$v_1 = \frac{2ga^2}{9\eta}(\sigma - \rho).$$

Everything in this equation is known or can be directly measured except a. Therefore from this equation, we can determine a and from it get the mass of the drop which will be $m = \frac{4}{3}\pi\sigma a^3$. The data to substitute in this equation and the additional data needed for the determination of e_i are collected in the following table:

$$\begin{aligned}
\eta &= 182.4 \times 10^{-7} \text{ newton-sec./sq. m.}\\
\sigma &= 919.9 \text{ kg./cu. m.}\\
\rho &= 1.2 \text{ kg./cu. m.}\\
\sigma - \rho &= 918.7 \text{ kg./cu. m.}\\
g &= 9.803 \text{ m./sec.}^2\\
v_1 &= 3.842 \times 10^{-4} \text{ m./sec.}\\
v_2' - v_2 &= 0.887 \times 10^{-4} \text{ m./sec.}\\
V &= 5{,}051 \text{ volts}\\
d &= 1.6 \times 10^{-2} \text{ m.}\\
F &= V/d = 5{,}051/1.6 \times 10^{-2} = 3.156 \times 10^5 \text{ volts/m.}
\end{aligned}$$

First, applying Stokes' law to get the radius of the drop, we obtain

$$a^2 = \frac{9\eta v_1}{2g(\sigma - \rho)} = \frac{9 \times 1.824 \times 10^{-5} \times 3.84 \times 10^{-4}}{2 \times 9.803 \times 918.7}$$

$$a^2 = \frac{7.004 \times 9 \times 10^{-9}}{2 \times 9.803 \times 918.7} = \frac{63.04 \times 10^{-9}}{2 \times 9006}$$

$$= \frac{63.04 \times 10^{-9}}{18.01 \times 10^3} = 3.500 \times 10^{-12}$$

$$a = 1.871 \times 10^{-6} \text{ m.}$$

$$a^3 = 6.55 \times 10^{-18} \text{ cu. m.}$$

Now, get the mass of the drop by multiplying the volume by the density.

$$m = \tfrac{4}{3}\pi a^3\sigma = 4.189 \times 6.55 \times 10^{-18} \times 919.9 = 2.524 \times 10^{-14}\,\text{kg.}$$

It is obvious from this figure why it is impossible to weigh the oil drop on a chemical balance. Using this value of m in Eq. (7) and taking the other data given above, we get

$$e_i = \frac{mg}{Fv_1}(v_2' - v_2) = \frac{2.524 \times 10^{-14} \times 9.803 \times 8.87 \times 10^{-5}}{3.156 \times 10^5 \times 3.842 \times 10^{-4}}$$

$$= \frac{2.195 \times 10^{-17}}{1.212 \times 10^2} = 1.81 \times 10^{-19}\,\text{coulomb.}$$

The preceding value is for the smallest change in the charge that was ever observed with the particular drop for which the data have been given. It is considerably larger than the value of 1.59×10^{-19} coulomb determined by Millikan by more careful experiments, corrected for various difficulties which we have not yet mentioned. (A redetermination of the viscosity of air alters the value of e obtained from Millikan's data to 1.60×10^{-19} coulomb.) The principal correction is for the failure of Stokes' law to hold exactly for drops as small as those used in these experiments. By studying many different drops, Millikan found that the value of e_i found for the smaller drops was consistently larger than that found for the larger drops. Now Stokes' law is derived on the assumption that the sphere is moving through a continuous medium, but we know that air is not continuous but is in reality made up of great numbers of tiny molecules with empty space in between them. As long as these

spaces are small compared with the size of the sphere, it is all right to think of the air as a continuous medium; but for drops as small as are being used here, this is no longer justifiable. In Chap. II, Par. 21, we saw that in oxygen at ordinary temperature and pressure the average distance between molecules was 3.3×10^{-7} cm. and that the total volume occupied by the molecules was only 1/2,000 of the whole volume of the gas. The volume of an oxygen molecule was given as 2.0×10^{-23} cc. We have just calculated the radius of one of our drops as 1.871×10^{-4} cm. and its volume as 2.74×10^{-11} cc. Thus the *average* distance between the air molecules through which the drop is moving is about $\frac{1}{500}$ of the radius of the drop. It is not surprising, therefore, that the use of Stokes' law for such a drop introduces considerable error. We shall not attempt to discuss how this error and other lesser ones were corrected. A full account of the procedure will be found in Millikan's book. We shall conclude our discussion of this classic experiment by giving the value that is now accepted as the most probable one for the charge on an electron. It is

$$4.80 \times 10^{10} \text{ e.s.u. of charge}$$

or

$$\mathbf{16.0 \times 10^{-20} \text{ coulomb.}}$$

SUMMARY

The atomic nature of matter suggests that electric charges may also be atomic, consisting of a large number of discrete particles of charge. Although the experiments we have been describing do not require this idea for their explanation, we have been using it when we spoke of electrons moving about in conductors. The evidence for the existence of electrons comes from several quarters. The first that we consider is that furnished by Millikan's oil-drop experiment.

Millikan observed the motion of charged droplets of oil in the region between the two plates of an electrical condenser. He first measured the rate of fall of a droplet under the action of gravity alone and then measured the rate of rise of the droplet when he applied an electric field that was more than sufficient to overcome gravity. In repeated observations, he found that the velocity of a given droplet always had certain definite values.

Simple equations show that the velocity upward must be always proportional to the charge on the droplet. Therefore, the charge always had certain definite values. Furthermore, these values were always integral multiples of one minimum value.

To determine the value of this minimum charge, it was necessary to know the mass of the oil droplet. Since this mass was far too small to measure by ordinary means, it was necessary to use Stokes' law for the rate of fall of a small sphere through a viscous fluid, namely $v = 2ga^2(\sigma - \rho)/9\eta$ where a is the radius of the sphere, σ its density, ρ the density of the fluid, and η its coefficient of viscosity. This law has to be modified for very minute droplets.

The conclusion drawn from Millikan's experiment brought up to date is that all electric charges both positive and negative are integral multiples of one minimum indivisible charge which has the value

$$16.0 \times 10^{-20} \text{ coulomb.}$$

ILLUSTRATIVE PROBLEMS

1. A Van de Graaff sphere stands isolated and charged to -10^5 volts. A charged pith ball with a mass of 80 mg. is held at a point of zero potential 51 cm. below the under side of the sphere. When released, the pith ball is attracted to the sphere and strikes it with a velocity of 1 m./sec. What is the charge on the pith ball? Neglect the viscosity of the air.

We shall work this problem from the energy principle. As the ball rises, the electrical potential energy decreases and the gravitational potential energy increases. The decrease in potential energy between the point of zero potential and the lower surface is thus the decrease of electrical minus the increase of gravitational potential energy. This decrease in total potential energy is equal to the increase of kinetic energy. The change of electrical potential energy is 0 volts $\times q - 10^5$ volts $\times q = -10^5\,q$ volt, a decrease. The change of gravitational potential energy is

$$\begin{aligned} mgh &= 80 \times 10^{-6} \text{ kg.} \times 9.8 \text{ m./sec.}^2 \times 0.51 \text{ m.} \\ &= 4.00 \times 10^{-4} \frac{\text{kg.-m.}}{\text{sec.}^2}\text{m.} = 4.00 \times 10^{-4} \text{ joule.} \end{aligned}$$

The decrease in potential energy is, therefore,

$$10^5\,q \text{ volts} - 4.00 \times 10^{-4} \text{ joule.}$$

The increase in kinetic energy is

$$\tfrac{1}{2}mv^2 = \tfrac{1}{2}80 \times 10^{-6}\text{ kg.} \times (1\text{ m./sec.})^2$$
$$= 40 \times 10^{-6}\frac{\text{kg.-m.}^2}{\text{sec.}^2} = 0.4 \times 10^{-4}\text{ joule}$$

since the initial kinetic energy is zero.

Equating the decrease in potential energy to the increase in kinetic energy, we obtain

$$10^5 q\text{ volts} - 4.00 \times 10^{-4}\text{ joule} = 0.4 \times 10^{-4}\text{ joule}$$
$$10^5 q\text{ volts} = 4.4 \times 10^{-4}\text{ joule}$$
$$q = \frac{4.4 \times 10^{-4}}{10^5}\frac{\text{joule}}{\text{volts}}$$
$$= 4.4 \times 10^{-9}\text{ coulomb.}$$

2. If a balloon of 10 cm. radius and an average density of 1.19 kg./m.^{3} rises in still air (density 1.20 kg./m.^{3}) with a velocity of 12 m./sec., what is the coefficient of viscosity of air?

Equation 3 (page 261) gives the terminal velocity of a sphere under the force of gravity. If the density σ is less than ρ, this velocity will be negative and hence up instead of down. The velocity is

$$v = \frac{2ga^2}{9\eta}(\sigma - \rho) \tag{3}$$

where $v = -12$ m./sec. = the terminal velocity.
$a = 10$ cm. = the radius of the sphere.
$g = 9.8$ m./sec.^{2} = the acceleration of gravity.
$\eta = ?$ = the coefficient of viscosity.
$\sigma = 1.19$ kg./m.^{3} = the density of the sphere.
$\rho = 1.20$ kg./m.^{3} = the density of the air.

Substituting these numerical values in Eq. (3) gives

$$-12\text{ m./sec.} = \frac{2 \times 9.8\text{ m./sec.}^2\,(0.1\text{ m.})^2}{9\eta}(1.19\text{ kg./m.}^3 - 1.20\text{ kg./m.}^3)$$
$$\eta = \frac{0.196\text{ m.}^3/\text{sec.}^2 \times 0.01\text{ kg./m.}^3}{9 \times 12\text{ m./sec.}}$$
$$= 1.82 \times 10^{-5}\frac{\text{kg.-m.-sec.}}{\text{m.}^2\text{-sec.}^2}$$
$$= 1.82 \times 10^{-5}\frac{\text{newton-sec.}}{\text{m.}^2}$$

PROBLEMS

1. Two large flat sheets of tin each of area 3 sq. m. are mounted parallel and 60 cm. apart. What is the capacity of this condenser? When it is charged to a potential difference of 50,000 volts with the positive plate uppermost, what is the total charge on one of the plates? Suppose a spherical balloon of 8 cm. radius with a conducting surface charged to a

potential of $-1{,}000$ volts is placed between the plates. What force will act on the balloon?

2. If the average density of the balloon in Prob. 1 is 1.28 kg./cu. m. and that of the air in the room 1.20 kg./cu. m., what is the resultant gravitational force on the balloon? What is the resultant of the gravitational and electric forces?

3. What potential would be required on the plates of the condenser of Prob. 1 to make the balloon of Probs. 1 and 2 stand still?

4. Two Van de Graaff spheres are placed one above the other with their centers 5 m. apart. The radius of the upper sphere is 50 cm. and its potential 10^5 volts. The radius of the lower sphere is 1 m. and its potential 10^6 volts. A 10-g. ball with a charge of 1 microcoulomb is released at the lower surface of the upper sphere. With what velocity will it strike the upper surface of the lower sphere? Neglect the viscosity of the air.

5. If the charge on an electron is 1.60×10^{-19} coulomb, the corresponding change in the velocity of rise due to the addition of this charge to a drop 0.3 mm./sec., and the potential between the plates of the Millikan condenser 5,000 volts; find the value of the terminal velocity of the drop under gravity.

6. If a spherical droplet of water falls at the rate of 1 m./sec. in still air, what is its radius?

7. What is the diameter of a ball bearing that falls with a speed of 1.69 cm./sec. in glycerin?

8. How much error in the terminal velocity of an oil drop of 1/1,000 mm. radius is made by neglecting the buoyant force of the air?

9. Express the data of Table B in the form of Table C.

10. How many elementary charges are there in the drop in each case of Table B?

11. If the values in Table C represent the velocities in hundredths of millimeters per second, what is the smallest change in velocity shown by the drop in Table C? If 0.522 cm. was the distance of travel, what is the change in the time of rise of the drop introduced by adding a single electron to the drop?

12. Let the values in Table C represent the velocities in hundredths of millimeters per second. If the terminal velocity of fall under gravity of the drop in Table C is 0.15 mm./sec., what will be the velocity of the drop with one elementary charge? What will be its minimum velocity upward, and how many elementary charges are required to produce this velocity?

13. With the same assumptions as in Prob. 12, between which pair of multiples of the elementary charge would the drop of Table B change its velocity from down to up?

14. With the same assumptions as in Prob. 12, calculate the value of the elementary charge from the data of Table C. Assume the same density of oil and air as in Table B, and the potential between the plates to be 1,620 volts.

CHAPTER XIII

BATTERIES, ELECTROLYSIS AND ELECTRIC CURRENTS

1. In the previous chapters, we have been concerned primarily with "static electricity," *i.e.*, with the effects of electric charges that were not moving or that were moving slowly. But we have indicated how electric charges are set in motion by an electric field, how positive charges tend to move from a place of high potential to one of lower potential; and, in Chap. XI, we stated that the setting up of potential differences was one of the most important problems of practical physics. In the present chapter, we shall give a brief statement of one practical way of setting up a potential difference and we shall then describe some electrochemical effects of the motion of charges and explain what is meant by an electric current.

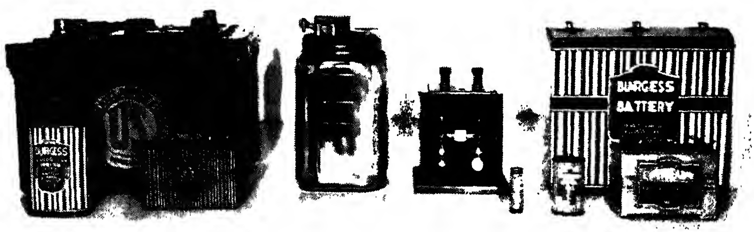


FIG. 107.—Various types of batteries.

2. In describing Millikan's experiment, we have spoken of the batteries used by him to produce a potential difference between the plates of a condenser. In the corresponding demonstration experiment, we used the Van de Graaff machine, which was understandable in terms of the simple experiments on static electricity. But batteries such as were used by Millikan form a much more satisfactory source of potential as well as a more familiar one. In general, a battery consists of a set of "cells" each one of which produces a certain definite potential difference.

There are many different types of such cells, but they all consist essentially of two conducting electrodes immersed in a chemical solution. Chemical reactions taking place between the material of the electrodes and the solution supply energy and maintain a difference of potential between the two electrodes, the exposed ends of which are called the terminals of the cell. There are many combinations of substances that give such effects. They can be divided into two classes, the reversible and the irreversible. The common storage battery is an example of the reversible type. If current is drawn from the cell, the chemical reactions proceed in one way; if current is driven back into the cell, the chemical reactions reverse. The ordinary "dry" cell is an example of the irreversible type. The details of the electrical and chemical effects that take part in producing the potential differences of these various cells are extremely complicated and none too well understood. The description given in the next paragraph does not pretend to be complete or exact. It is included with some hesitation in the hope that it may interest the student in this important problem and help him to understand the later paragraphs on electrolysis.

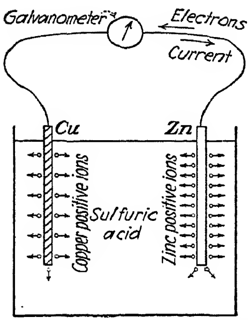


FIG. 108.—Potential difference set up by a simple copper-zinc-acid cell.

A Particular Cell

3. If a plate of copper and a plate of zinc are dipped into a dilute solution of sulfuric acid, it is found that a difference of potential exists between the exposed ends. This is interpreted as follows: At the surface of the copper electrode, atoms of copper are continually being dissolved by the action of the acid. However, they go into solution not as neutral atoms but as positively charged atoms, or ions as they are called, leaving electrons behind them on the copper. This sets up a difference of potential at the surface of the copper, the copper becoming negative with respect to the solution. The same thing happens at the surface of the zinc where positively charged zinc ions go into solution, leaving the zinc plate negatively charged and at a lower potential than the solution. But the tendency of the zinc to go

into solution is greater than that of the copper so that a greater difference of potential is set up between the zinc and the solution than between the copper and the solution. Since both are at negative potentials with respect to the solution, this means that the copper is less negative than the zinc, *i.e.*, is positive with respect to the zinc. Consequently, if the exposed ends of the plates are connected by a conductor, current will flow, electrons being drawn from the zinc to the copper. In the solution, on the other hand, the positive zinc ions are going into solution more rapidly than the positive coppers at the other electrode, in fact, it is this that produces the difference of potential between the zinc and the copper, so that there is an excess of positives built up in the solution in the neighborhood of the zinc. These tend to repel each other out into the solution to such an extent that there is a resultant flow of positive charge into the copper plate in spite of the copper ions continually coming into solution. This flow of positives just counteracts the flow of electrons through the wire so that no extra resultant charge builds up on the electrode. If the wire between the two electrodes is disconnected, the potential differences at the electrodes build up to a sufficient strength to slow up the chemical reaction, but the ions still diffuse out through the solution under the influence of the thermal agitation and the formation of the sulfates continues, the energy released going into heat instead of the motion of electrical charges. The details of the chemical reactions in the cell are involved, but one effect that is important is the formation of hydrogen at the electrode which reduces the potential difference between zinc and copper. This "polarization," as it is called, makes this type of cell unsatisfactory in practice. Batteries can be made that avoid this difficulty, but we shall leave the complexities of this topic to physical chemistry.

Any Cell

4. From our point of view, the details of this and similar processes do not have to be considered. We need only know that it is possible by the use of certain combinations of substances to set up potential differences that can be used to do work and that the energy used in doing the work is supplied by chemical reactions much as the energy used by a steam engine comes from the chemical reactions in the burning of the fuel. Furthermore, the

potential differences set up by a particular combination of substances is always the same. It is a familiar fact that the ordinary "dry" cell gives about 1.5 volts, and each cell of a storage battery gives about 2 volts. These are small potential differences compared with those that we get by frictional electricity, induction machines, or the Van de Graaff machine. But they have the great advantage of dependability, reproducibility and, most important of all, they can supply much larger amounts of energy than the devices just mentioned can supply. They are like a great river which can supply much greater energy at a dam 20 ft. high than can be obtained from a small stream falling over a precipice 1,000 ft. high.

Electromotive Force

5. The potential differences produced by batteries are usually called electromotive forces or e.m.fs. In general, the term electromotive force is used for potential differences set up by agencies such as chemical reactions, the motion of the parts of a dynamo, the heating of a thermojunction, and the effect of changing magnetic fields. As the words imply, it is always an effect that tends to move electric charges. In the absence of an e.m.f., the currents in a circuit will gradually die out and every point in the circuit will come to the same potential. To maintain a steady flow of current, an e.m.f. must be present and be continually supplying energy.

The Daniell Cell. The Volt

6. A type of battery used almost universally in the early days of telegraphy as a source of e.m.f. was that invented by Daniell. It is essentially like that described in Par. 3 but with the copper electrode submerged in copper sulfate which is separated from the dilute sulfuric acid (or zinc sulfate) around the zinc by a porous partition. The hydrogen formed by the action of the cell replaces the copper in the copper sulfate. The displaced copper instead of the hydrogen then goes to the electrode, and polarization is thus avoided. Because of the widespread use of this cell and its convenient size, it came to be generally accepted as a unit of e.m.f. The e.m.f. of a Daniell cell is approximately one volt but, for theoretical reasons, the precisely defined volt, the "practical" unit of e.m.f., is slightly different. For our purposes, it is

best expressed in terms of a cell that gives exceptionally reproducible results, namely, the cadmium standard cell, also called the Weston standard cell after its inventor. The volt is then defined by the following statement:

The electromotive force of the cadmium standard cell at 20°C. *shall be taken as equal to* 1.0179 *volts.**

Batteries in Series and in Parallel

7. The two standard ways of connecting a number of batteries are probably familiar but are of sufficient importance to be

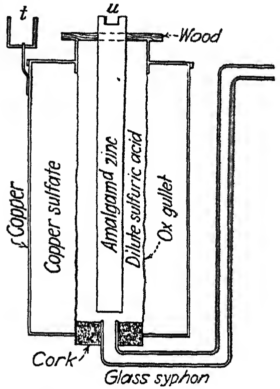


FIG. 109(*a*).—Original Daniell cell. The ox gullet served the purpose of a permeable partition corresponding to the earthenware cup used later. Connections were made through mercury cups at *t* and *u*.

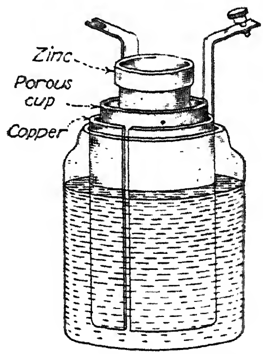


FIG. 109(*b*) —Daniell cell. The space inside the cup is filled with dilute sulfuric acid. The outer chamber contains a solution of copper sulfate.

recalled. In one method, the batteries are connected "in series" as shown in Fig. 110(*a*), *i.e.*, the positive terminal of each battery is connected with the negative terminal of the next one. It is obvious from the definition of potential difference that the poten-

* Though we are using the Weston cell to define the volt, we are defining the "absolute volt" not the "international volt" because of the recent agreement of the International Committee on Weights and Measures to discard the "international" units after Jan. 1, 1940. Strictly speaking, the absolute volt stems from the ampere and the joule, but we are using the preceding definition because of its greater simplicity for the student.

tial difference between the positive terminal of the first battery and the negative terminal of the last is the sum of the potential differences across the separate batteries, if the potential drop in the connecting wires is negligible. Such a method, therefore, gives a large e.m.f. But the current that can be supplied by the set of batteries is limited by that which one can supply. If, on the other hand, all the positive terminals are connected together and all the negative terminals are connected together, the batteries are said to be connected in parallel. The current that can be drawn from such a group of batteries is equal to the sum of what can be drawn from each separate battery, but the e.m.f. supplied is equal to the e.m.f. of one.

Electrolysis

8. Suppose that the two electrodes of the cell described in Par. 3 had not been connected together by a wire but had been connected to the terminals of a powerful battery giving an e.m.f. overwhelmingly larger than that of the single cell. Then the natural differences of potential at the electrodes would be small compared with the difference of potential imposed by the large battery. If the positive terminal of the large battery were connected to the copper electrode of the cell and the negative to the zinc electrode, the positive ions of the copper in solution would be repelled by the positive copper electrode and pulled strongly toward the negative zinc. The positive zinc ions would never get away from the strong attraction of the negative zinc electrode. Consequently, there would be a steady flow of copper ions toward the zinc electrode where they would tend to give up their positive charge and deposit on the electrode. The zinc would thus become plated with copper. If the connections were reversed so that the zinc electrode were connected to the positive terminal of the big battery, the flow of ions would be reversed so that the zinc would be deposited on the copper.

9. This same process may occur in any solution that contains ions, *i.e.*, charged molecules, whether the materials be such as to produce a natural e.m.f. in the cell or not. For example, both electrodes might be made of the same material so that there was no difference of potential between the two electrodes until they were connected to some outside source of e.m.f. The imposed e.m.f. would then merely cause the transfer of the molecules of

one electrode to the other. In some cases, the transfer of ions from one electrode to the other is accompanied by chemical reactions at the surfaces of the electrodes. This is the case in the familiar example of the electrolysis of water acidulated by a little sulfuric acid, where the final results of the electrolysis are hydrogen gas at the negative electrode and oxygen gas at the positive electrode. The details of this and similar processes are still somewhat controversial and need not concern us here. We shall confine ourselves to a statement of the general laws that Faraday found governing all the effects of the passage of electric current through solutions. Faraday's Laws of Electrolysis areas follows:

1. *When varying quantities of electricity pass through the same solution, the quantities of materials set free at either electrode are proportional to the quantities of electricity that have traversed the solution.*

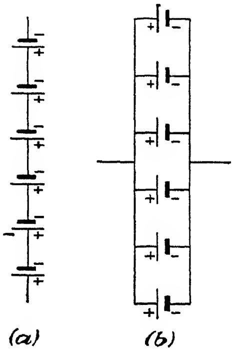


FIG. 110.—(*a*) n batteries connected in series. E.m.f. = n times that of one battery. (*b*) n batteries connected in parallel. Current capacity = n times that of one battery.

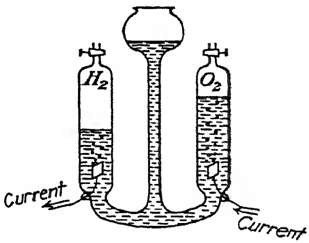


FIG. 111.—Electrolysis of acidulated water. Since an ion of oxygen carries twice as much charge as one of hydrogen, the volume of oxygen set free by a given charge will be half that of the hydrogen.

2. *When the same quantity of electricity passes through several different solutions, the weights of the various materials liberated are directly proportional to their chemical equivalents.*

(The chemical equivalent of a chemical element is the atomic weight divided by the valence, which is itself always a small whole number. Thus, the chemical equivalent is either equal to the atomic weight or to the atomic weight divided by two or three, or sometimes a higher number.)

Interpretation of Faraday's Laws

10. In the first chapter, we saw that all substances are made up of atoms so that setting free a certain number of grams of an element, copper, for instance, at an electrode corresponds to setting free a certain number of atoms of copper at that electrode. In the last chapter, we saw that all electric charges were made up of atoms of electric charge so that the passage of a certain quantity of electricity through a solution corresponds to the passage of a certain number of electronic charges through that solution. Thus, according to Faraday's first law, the number of atoms deposited is always proportional to the number of electronic charges that have passed through the solution as long as we are dealing with the same substances. The simplest explanation that we can suggest for this parallelism is that one electronic charge is associated with each atom. Let us consider the second of Faraday's laws and see whether it supports this suggestion.

11. At first, let us confine ourselves to monovalent substances. Then Faraday's second law states that the amount of an element deposited by a given amount of electric charge, *i.e.*, a given number of electronic charges, is proportional to the atomic weight. But we saw in the first chapter that the weights of the individual atoms of an element are proportional to the atomic weight of that element. Faraday's second law, therefore, means that the same number of atoms are deposited by a given number of electronic charges for all monovalent elements. Clearly, this is in accord with the suggestion that one electronic charge is associated with each of the atoms deposited on the electrode, confirming the deduction from the first law.

12. To explain the results with elements of valence greater than one, it is necessary to add only the assumption that an atom can and does carry a number of electronic charges equal to its valence. Such elements are of great importance in chemistry but need not concern us further.

Avogadro's Number

13. In the early chapters of this book, we defined Avogadro's number as the number of molecules in a mole or gram molecule of any substance. For a monovalent element, this is also the number of atoms in a chemical equivalent of that element.

Therefore, Avogadro's number is the number of atoms that have passed through the solution in a process of electrolysis when one chemical equivalent has been deposited. We can measure the charge that has passed through the solution in the same process. According to Faraday's second law, it will be the same for all substances and it is found experimentally to be equal to 9.65×10^4 coulombs. Assuming that each atom carries a single electronic charge, we can get the total number of atoms that have passed by dividing the total charge by the charge on each atom, *i.e.*, by 1.60×10^{-19} coulomb, the electronic charge determined by Millikan's experiment. When the values just given are used, this calculation gives for Avogadro's number the figure 6.03×10^{23}. This is the value for Avogadro's number that was quoted in Par. 37 of Chap. I, and this determination by combining the data of electrolysis with the results of Millikan's experiment is probably the most accurate way of determining Avogadro's number. That it agrees within the limits of error with the other methods may be taken as confirmation of the theory of electrolysis on which it is based or of Millikan's results. Before the latter were available, it was possible to calculate the electronic charge by taking some other determination of Avogadro's number and reversing the preceding calculation.

The Unit of Charge

14. In Chap. X, we defined unit quantity of electric charge in terms of Coulomb's law by using a factor of proportion that had an arbitrary value. We are now in a position to justify the choice of that particular value. From our definition of potential difference, we saw that the product of a charge and a potential difference was energy. Since we have the joule as a unit of energy from mechanics and have just defined the volt as the unit of e.m.f. or potential difference, we can define a unit of charge in terms of them as follows:

One coulomb is the amount of electric charge that requires one joule of energy to move it through an opposing potential difference of one volt.

The value of the constant of proportion given in Eq. (2), Par. 13, Chap. X, was so chosen as to be consistent with the foregoing definition which is really what determines the value of the coulomb.

Measurement of Charge and Current by Electrolysis

15. Evidently, it is possible to determine the number of coulombs of charge that have passed through an electrolytic cell by determining the amount of material carried to the electrodes. Furthermore, if the time is measured, the average rate of flow of charge, *i.e.*, the average current, can be determined. This method was formerly much used because of its simplicity and accuracy. The coulomb was even defined as the amount of charge that would deposit 0.001118 g. of silver from solution. These methods have been almost entirely superseded by others depending on the magnetic effects of currents which will be described in later chapters.

Electric Current. The Ampere

16. The phenomena we have been discussing depend on the total amount of electric charge that flows through a solution, not

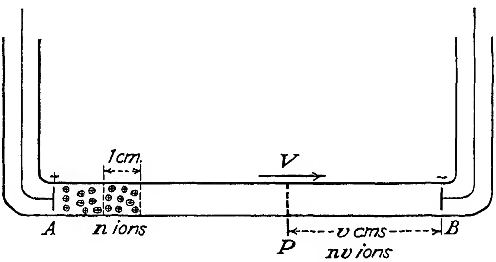


FIG. 112.—Motion of ions in an electrolyte. The tube has a cross section of 1 sq. cm. and there are n ions per cubic centimeter, each moving on the average with a velocity v. (The ions are indicated in only one part of the tube. Their distribution is assumed uniform.)

on the rate at which it flows. The amount of silver deposited on an electrode by a flow of a million atoms per second for a period of 1 hr. is the same as the amount of silver deposited by a flow of two million atoms per second for $\frac{1}{2}$ hr. Similarly, the amount of charge carried by one million ions per second flowing for 1 hr. is the same as that carried by two million ions per second for $\frac{1}{2}$ hr. Nevertheless, there are other effects that make it desirable to study this question of rate of flow in detail.

17. Let us consider an electrolytic cell such as is shown in Fig. 112 where the two electrodes are connected by a glass tube of

1 sq. cm. cross section and full of solution. Suppose that the only ions in the solution are atoms of silver each with a single positive electronic charge. The electrode A is positive and the electrode B is negative so that the ions move from A to B under the influence of the electric field. Their motion will be hindered by the random thermal motion of the molecules which they will be continually hitting, but we may assume that they all have a resultant motion from A toward B with a velocity v. Let the mass of each ion be m, its charge be e, and let n be the number present in each cubic centimeter of the solution. (Of course, for the specific case we have chosen, m is $107.88/6.03 \times 10^{23}$ g. and e is 1.60×10^{-19} coulomb.) Take any instant of time, and consider all the ions that are v cm. or less away from B. In the course of the next second, all these ions will get to B since they are moving with a velocity of v cm./sec. toward B and are at most only v cm. away. No ions that were farther to the left at the beginning of the second will quite get to B. Since there are n ions/cc. by definition and since the tube is 1 sq. cm. in cross section, the number of ions that get to B per second will be nv. By the same argument, this will be the number per second flowing through any plane perpendicular to the tube. The mass of silver carried by these ions will be nvm g./sec., and the electric charge they carry will be nve units of charge per second. Thus we see that we can express the flow of ions to the electrode by saying that there are so-and-so many ions per second reaching it, that there is such-and-such a number of grams of silver being deposited there per second, or that there are so-and-so many coulombs of electricity getting there per second. Any one of these statements might describe a property of the electrolysis that we might be interested in. At the moment, we are most concerned with the last, the rate at which electric charge is being carried to the electrode. We say that there is an electric current flowing, and we find that we have to deal with such currents so frequently that we give names of their own to the units of electric current. *Thus the flow of one coulomb of electric charge per second is called a current of one ampere.*

18. In the example of electric current discussed in the previous paragraph, the current was carried by the transfer of charged particles set in motion by an electric field. We now believe that all electric currents are carried in this way. In the conduction of

current through a gas, much of the charge is carried by the motion of charged atoms as in electrolysis but, in the gas, some of the current is also carried by still smaller particles of matter, electrons, each of which carries one negative electronic charge and is about 1/1,800 of the mass of a hydrogen ion. The motion of negatively charged electrons in one direction is the equivalent of a motion of positive charges in the other direction so that to get the total current passing through a gas the contributions of the two different carriers have to be added. In solid conductors, the atoms are so closely packed and tightly bound that they are not able to move so that any electric current flowing in solid conductors has to be carried by electrons. That this is the case has already been assumed in our earlier discussion of conductors.

Thermoelectric Effect

19. In a battery, the forces at the surfaces separating solids and solutions are used to get a resultant e.m.f. Between two metals in contact, there are similar forces that can be used to get measurable though small e.m.fs. The amount of these differences of potential depends on temperature; and if we deal with a circuit containing two identical surfaces of contact at different temperatures, appreciable e.m.fs. are obtained.

20. The first recorded observations of this effect were described by Seebeck in 1822. His circuit is shown in Fig. 113. It consisted of a galvanometer in series with two pieces of copper strip whose ends A and B were joined by a piece of bismuth. He found that the galvanometer showed a flow of current in one direction if the junction A was hotter than the junction B and that the current reversed in direction if B was made hotter than A.

21. Experiments since Seebeck's time have shown that this is a perfectly general result. Any two metals or alloys connected in a circuit like that of **Fig. 113** will set up a current in the circuit if the junctions are at different temperatures. This effect is called *the thermoelectric effect* or sometimes the Seebeck effect. It is found that the amount of resultant e.m.f. developed by the two junctions depends on the nature of the metals and the temperature of the junctions. For some pairs of metals, over considerable ranges of temperature, the e.m.f. developed is directly proportional to the difference of temperature between the two junctions. But this is not true over the whole range of tempera-

ture. In fact, the behavior of iron and copper is quite typical. For instance, if a circuit like that of Fig. 114 is made with iron replacing bismuth and one junction kept in an ice bath at zero degrees while the other junction is heated, the galvanometer

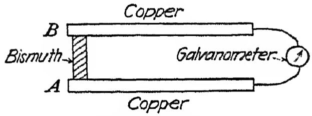


FIG. 113.—Thermoelectric circuit used by Seebeck. If the junctions A and B are at different temperatures, a current will flow through the galvanometer.

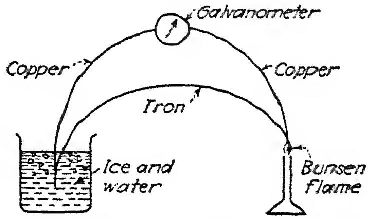


FIG. 114.—Thermoelectric circuit with two iron-copper junctions, one at 0°C. and one at a higher temperature.

deflection will increase up to a certain temperature of the hot junction, then decrease to zero, and then reverse at still higher temperatures. A graph of the results for iron-copper junctions is given in Fig. 115. In contrast to the iron-copper curve is the

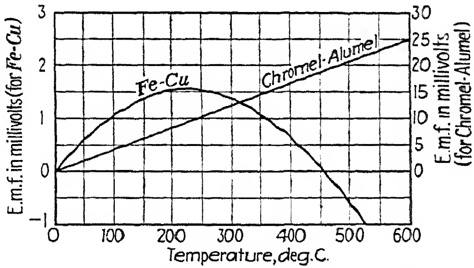


FIG. 115.—E.m.fs. of iron-copper and chromel-alumel thermocouple circuits as functions of the temperature of the hot junction. The cold junction is at 0°C.

straight line (Fig. 115) for chromel-alumel, two alloys frequently used in practical temperature measurement.

22. The e.m.fs. developed in the thermoelectric circuits are so small that they are practically useless as sources of power. Their importance is in the measurement of temperature. A combination of wires of two kinds designed to measure temperature is called a thermocouple. Thermocouples can be used over

much greater ranges of temperature than glass thermometers since they are limited only by the melting points of the metals. Also they can be made very much smaller than any glass or metal expansion thermometer since the e.m.f. developed depends only on the nature of the wires not on their size. Consequently, they can be used to measure the temperature of very small objects such as insects; or the rise in temperature in a very small thermocouple can be used to measure a small quantity of energy. Thermocouples are also well suited to the measurement of temperatures in inaccessible places since it is only necessary to lead out two fine wires from the point where the temperature is being measured to the cold junction and measuring instrument.

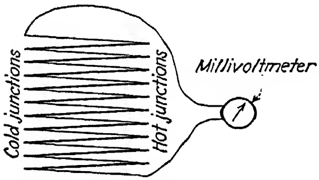


FIG. 116.—Schematic diagram of a ten-element thermopile.

23. For some purposes, it is useful to have a group of thermocouples connected in series with alternate hot and cold junctions as shown in Fig. 116. Such an instrument is called a thermopile and can be made sensitive to very small quantities of heat. For example, it can be used to measure the heat coming from a star.

24. The values of the thermal e.m.fs. developed between the hot and cold junctions of various thermocouples for various differences of temperature are given in Table 13. Many of the materials used in thermocouples are alloys since it has been possible to get larger e.m.fs. and ones that vary more uniformly with temperature by using alloys than by using pure metals. As can be seen from the table, the e.m.fs. developed are of the order of magnitude of a hundredth of a millivolt per degree difference in temperature between the junctions.

25. We shall attempt to give a crude explanation of the thermoelectric effect in terms of the electrical structure of the metals concerned. In order to be specific, let us go back to the case of iron and copper. Suppose that in both metals there are electrons moving among the atoms under the influence of thermal agitation. Imagine a plane through the copper at any point. There must be a number of electrons crossing this plane every second, some going in one direction some in the other. But since the structure of the metal, the spacing of the copper atoms, and the forces with which they attract electrons are the same on each

side of the plane, the average number of electrons crossing the plane in one direction must be just equal to the number crossing in the other direction so that there is no accumulation of charge

TABLE 13.—THERMOELECTRIC E.M.FS

Values are given in millivolts for several pairs of metals with the hot junction at a temperature t°C. and the cold junction at 0°C.

	Ag-Cu	Ni-Cu	Fe-Cu	Ag-Pt	Chromel-alumel
100	-0.04	- 2.26	+1.06	+0.72	4.08
200	-0.009	- 4.89	+1.52	+1.73	8.17
300	-0.17	- 7.56	+1.38	+2.96	12.32
400	-0.17	- 9.74	+0.64	+4.47	16.50
500	-0.13	-11.94	-0.70	+6.26	20.76

The plus sign indicates current flowing from the hot to the cold junction in the first metal of the pair.

on either side of the plane. Now consider the surface of separation between the copper and the iron. Here the structure is no longer identical on different sides of the surface. The electrons may be able to move more freely in the copper than in the iron, or vice versa. The "vapor pressure" of electrons in one direction may be greater than in the other. Specifically, in an iron-copper junction below the neutral point, more electrons tend to go from the iron to the copper than vice versa. In other words, the difference in structure between the iron and the copper produces electrical forces that move the electrons just as they would be moved by a tiny battery whose positive pole was connected to the iron and negative pole to the copper. We can describe the effect, therefore, by saying that there is an e.m.f. developed at the junction.

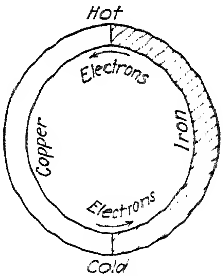


FIG. 117.—Enlarged schematic drawing of an iron-copper thermocouple circuit showing the direction of motion of the electrons.

26. We have described the effect at one junction, but the complete circuit as shown in Fig. 117 has two junctions, and it is evident that the e.m.f. at the second junction will be in the opposite direction to that at the first. If the e.m.f.s. are of the same size, there will be no resultant e.m.f. in the circuit and

no current will flow. But the motion of the electrons in the metal may be expected to depend on the temperature just as the motion of the atoms in a gas or solid depends on the temperature. Therefore, if the temperature of one junction is different from that of the other, the e.m.f. may be expected to be different at one junction from that at the other so that there will be a resultant e.m.f. in the circuit and current will flow. This is, of course, the observed result. The change in direction of the resultant e.m.f. at higher temperatures is to be interpreted in terms of progressive changes in the arrangement of the copper and iron atoms as the temperature rises. Any attempt to follow this in detail is quite beyond the scope of this book.

27. So far, we have omitted mention of one fact of great practical importance, the effect of introducing other metals in the circuit. If some other metal is put in between the copper and iron, aluminum, for example, the resultant e.m.f. in the circuit is unchanged so long as the copper-aluminum and iron-aluminum junctions are at the same temperature. This is a general result for any metals anywhere in the circuit and can be proved by a thermodynamic argument as well as experimentally. A particular case of it is one where the circuit is broken for the inclusion of a galvanometer. In this case, the circuit would probably be broken by cutting the copper wire, leaving the two iron-copper junctions unaffected. The new junctions between the copper and the terminals of the galvanometer and inside the galvanometer would then all be at room temperature and would give no resultant e.m.f.

28. Before going on to a discussion of some of the properties of electric currents, we shall list the different concepts that we have introduced so far in the study of electricity and warn the student that he must be clear as to their meaning before he can hope to understand what is to follow. They are

Electric charge or quantity of electricity.
Electric field strength or electric intensity.
Electric difference of potential. Electromotive force.
Electric capacity.
Electric current.

SUMMARY

Chemical energy is one of the most valuable sources of electrical energy, and chemical reactions are very generally used to

produce the potential differences required for electrical effects. The common storage battery and "dry cell" are both examples of such sources of electrical potential difference. A very simple type of battery consists of a plate of zinc and a plate of copper immersed in sulfuric acid. A difference of potential is set up between the zinc and copper, and current will flow from the copper to the zinc if they are connected by a wire. The advantages of batteries over electrostatic generators are greater compactness and reliability and, above all, greater energy capacity. The differences of potential produced are rather low, about one or two volts per cell, depending on the substances reacting. Differences of potential produced by cells or other outside agencies in a circuit are called electromotive forces. One particular type of battery, the Daniell cell, was used so generally in telegraphy that its e.m.f. of about one volt came to be taken as standard. A more reliable definition of the volt is in terms of the e.m.f. of a Weston standard cell taken as 1.0179 volts.

When batteries are connected in series, the resultant e.m.f. is the sum of the e.m.fs. of the separate batteries. When batteries are connected in parallel, the resultant e.m.f. is that of a single battery (assuming them all the same).

Just as chemical reaction can cause current to flow, a forced flow of current through a solution causes decomposition. This process is called electrolysis. The products of decomposition are set free at the electrodes. The laws governing electrolysis were studied by Faraday, who found that

1. When varying quantities of electricity pass through the same solution, the quantities of materials set free at either electrode are proportional to the quantities of electricity that have traversed the solution.

2. When the same quantity of electricity passes through several different solutions, the weights of the various materials liberated are directly proportional to their chemical equivalents.

These laws suggest that the current is carried by individual molecules each carrying one or more electronic charges. On this basis, the results of electrolysis give a value of Avogadro's number equal to 6.03×10^{23}.

The coulomb is defined as the charge that requires one joule of energy to move it through an opposing potential difference of one volt. An ampere is a current of one coulomb per second. Current is the motion of large numbers of individual charges.

Thermoelectric e.m.fs. are produced in a circuit when it contains two different metals whose two junctions with each other are at different temperatures. The amount of e.m.f. produced depends only on the nature of the metals and the temperatures of the junctions. Thermocouples use this effect to measure very extreme temperatures or the temperatures of inaccessible points.

The electrical concepts introduced up to this point are as follows:

Electric charge or quantity of electricity.
Electric field strength or intensity.
Electric difference of potential. Electromotive force.
Electric capacity.
Electric current.

ILLUSTRATIVE PROBLEMS

1. How many grams of copper will be deposited from a solution of copper sulfate by a current of 10 amp. flowing for $\frac{1}{2}$ hr.?

The mass of any substance set free from solution is proportional by Faraday's first law of electrolysis to the charge that has passed through the solution. Since the charge is the product of the current and the time during which it exists,

$$\begin{aligned} q &= it = 10 \text{ amp.} \times \tfrac{1}{2} \text{ hr.} \times 60 \text{ sec./hr.} \\ &= 300 \text{ coulombs.} \end{aligned}$$

The mass of copper liberated is, therefore, proportional to 300 coulombs, and the constant of proportionality is known as the electrochemical equivalent of copper. Thus the electrochemical equivalent is the number of grams liberated per coulomb. We have seen in Chap. I that there are 6.03×10^{23} atoms in 1 gram atomic weight of any substance, in this problem in 63.6 g. of copper. Since the valence of copper is 2, each of these atoms in solution will carry 2 positive charges, each positive charge equal to the charge on an electron or 1.60×10^{-19} coulomb. Therefore the charge required to deposit 6.03×10^{23} atoms of copper, or 63.6 g., is $6.03 \times 10^{23} \times 2 \times 1.60 \times 10^{-19}$ coulomb. The number of grams deposited by 1 coulomb is, therefore,

$$\frac{63.6 \text{ g.}}{6.03 \times 10^{23} \times 2 \times 1.60 \times 10^{-19} \text{ coulomb}} = 3.30 \times 10^{-4} \text{ g./coulomb.}$$

This is the electrochemical equivalent of copper. The mass of copper liberated by 300 coulombs is therefore

$$m = 3.30 \times 10^{-4} \text{ g./coulomb} \times 300 \text{ coulombs} = 0.099 \text{ g.}$$

2. If 300 coulombs are passed through a sodium chloride solution, how much sodium will be liberated? How much chlorine?

Since the masses of different substances liberated by the same charge are proportional to their chemical equivalents, we have, from Prob. 1, where 0.099 g. of copper is liberated by 300 coulombs,

$$\frac{0.099 \text{ g.}}{63.6 \text{ g.}/2} = \frac{m_{Na}}{23 \text{ g.}/1} = \frac{m_{Cl}}{35.46 \text{ g.}/1}.$$

m_{Na} is the mass of sodium and m_{Cl} the mass of chlorine liberated by 300 coulombs. The fractions in the denominator are the chemical equivalents, *i.e.*, the atomic weights divided by the valence, of copper, sodium, and chlorine, respectively.

$$m_{Na} = \frac{2 \times 0.099 \times 23 \text{ g.}}{63.6} = 0.0716 \text{ g.}$$

$$m_{Cl} = \frac{2 \times 0.099 \times 35.46 \text{ g.}}{63.6} = 0.110 \text{ g.}$$

PROBLEMS

1. Three identical cells each have an e.m.f. of 1 volt. What is the e.m.f. of (*a*) the three cells in series, (*b*) the three cells in parallel, (*c*) two cells in parallel in series with the third? What fraction of the current in the circuit flows through each cell?

2. How many grams of silver will be deposited from a solution of silver nitrate by a current of 5 amp. flowing for 1 hr.?

3. How many coulombs are required to deposit the chemical equivalent of silver? This is known as the Faraday. Show that it is the same for all elements.

4. Calculate the current required to deposit 10 g. of copper per hour in any electrolytic cell containing copper sulfate solution, the Faraday constant being given as 96,485 coulombs per chemical equivalent.

5. A current deposits 5 g. of silver from a silver nitrate solution. How many grams of copper will it deposit from a copper sulfate solution?

6. Calculate the electrochemical equivalent of nickel?

7. Compute Avogadro's number from an experiment in electrolysis.

8. A current of 15 amp. deposits zinc from zinc sulfate. How many grams of zinc are deposited per second? How many atoms?

9. If 5 amp. are depositing copper from copper sulfate, how many ions per second reach the electrode?

10. An electrode has an area of 100 sq. cm. If it is placed in a copper-plating bath and a current of 1 amp. flows for 1 hr., how thick is the deposit of copper?

11. A current of 10 amp. liberates hydrogen from acidulated water. How many cubic centimeters of hydrogen under normal conditions of temperature and pressure are liberated per second?

12. A current of 3 amp. liberates 690 cc. of hydrogen in ½ hr. at a pressure of 75 cm. and temperature of 20°C. What is the electrochemical equivalent of hydrogen? What volume of oxygen would be liberated simultaneously? What is the electrochemical equivalent of oxygen?

13. The hot junction of a chromel-alumel thermocouple is in a bath of molten lead and the cold junction is in ice water. A voltmeter in the circuit reads 0.02 volt. What is the temperature of the lead?

14. A circuit containing four silver-platinum thermojunctions in series has a resistance of 0.5 ohm. If alternate junctions are in ice water and in boiling water, what current is flowing in the circuit? (For Ohm's law see Chap. XVII.)

CHAPTER XIV

FORCES BETWEEN CHARGES IN MOTION

1. In the last chapter we defined electric current and discussed the simple example of the current carried by ions moving through a solution. The motion of these ions was caused by the electrostatic field between the electrodes immersed in the solution. But this is not the only kind of force that maintains electric currents; nor are the ions of atomic or molecular size the only carriers of electric charge that are important in the flow of currents. In the present chapter we shall discuss another, smaller carrier of electric charge and the forces set up by the motion of charges.

Discharge in Gases at Low Pressure. Cathode Rays

2. We found that the Brownian motion in gases gave us one of the most direct proofs of the existence of atoms and that it was in gases that we could most easily apply the atomic point of view to the actual behavior of large amounts of material. Now let us turn to the passage of electricity through gases to look for evidence about the nature of electric current. Let us study the discharge of electricity in air at reduced pressure. Suppose that a long straight glass tube with two metal electrodes (Fig. 118) is connected to a source of high potential (5,000 to 10,000 volts), such as an induction coil or a set of batteries, and that the air in the tube is gradually pumped out. At atmospheric pressure, nothing happens. Then as the pressure is reduced, an irregular pencil-shaped spark begins to connect the two electrodes. As the pressure becomes lower, this becomes a steady brightly colored discharge filling the whole space between the electrodes. As the pressure gets still lower, a dark space begins to develop in front of the cathode (the electrode connected to the negative

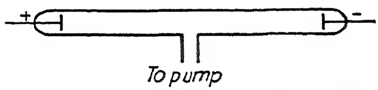


FIG. 118.—A tube for discharge in gases at low pressure.

side of the source of potential). As the pressure is still further reduced, this dark space fills the whole tube except for a beam of rays coming out from the cathode in a straight line. Finally, when the tube is very highly evacuated, the discharge disappears entirely. The rays that appear emerging from the cathode in the last stage of the discharge are called cathode rays.

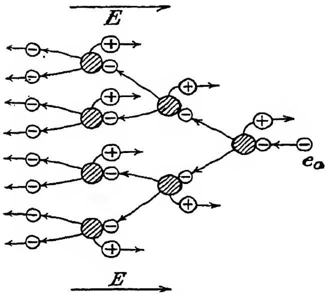


FIG. 119.—Detail of ionization in a discharge. Schematic drawing of cumulative ionization by electron impact. Starting with a single electron e_0 moving to the left under the action of the field E, each collision produces an additional electron and a positive ion. By the time the original electron has made three ionizing collisions, seven additional electrons and seven positive ions have been formed.

3. These phenomena are qualitatively understandable in terms of electrons and atoms. The potential source produces a strong electric field between the electrodes so that any ion in this space is acted on by a force and tends to be accelerated. There are always some ions present, produced by cosmic rays which knock negative electrons out of the gas molecules, leaving them positively charged. These few electrons and positively charged molecules tend to speed up under the action of the electric field, but, at high pressures, they keep hitting the uncharged air molecules so often that they can never get up much speed. As the pressure is reduced, the ions can go farther between collisions and so can acquire greater velocity. Soon they get so much kinetic energy that when they strike uncharged molecules they knock more electrons out of them, making more pairs of ions (Fig. 119). These, in turn, are accelerated by the electric field and produce more ions until there are soon enough to carry a comparatively heavy current (perhaps 20 to 50 milliamp.) and produce luminosity. To explain the nonuniformity of this luminosity is a more complicated matter, decidedly beyond the scope of this book. We can only say that at low pressure almost the whole of the potential drop is concentrated near the cathode. This means that the force acting on electrons set free near the surface of the cathode is very great so that they shoot out perpendicularly from this surface with enormous velocities, perhaps 10^5 or 10^6 m./sec. It is electrons shot out in this way that

form the cathode rays appearing at the lowest pressure before the discharge goes out. The final disappearance of the discharge results when the gas has been so completely removed that there are comparatively few molecules left to be ionized and, therefore, too few ions to carry the current.

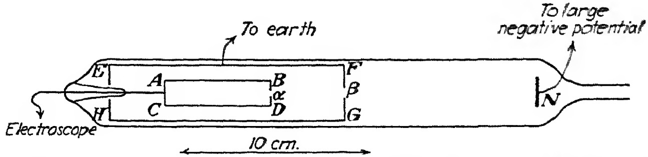


FIG. 120.—Perrin's experiment. Cathode rays from N pass through the hole β in the anode and are trapped in the insulated cylinder $ABCD$ that is connected to an electroscope. If the cathode rays are deflected from β by a magnet, the electroscope does not become charged.

4. It is the study of gas discharges in general and of cathode rays in particular that has been largely responsible for the development of the physics of the past forty years. We shall keep coming back to it. At present, let us consider some of the properties of cathode rays. That they travel in straight lines, cast shadows, cause glass and other substances to fluoresce brightly, heat up anything they strike, and so on can be shown

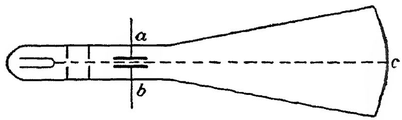


FIG. 121.—Cathode-ray oscillograph tube. The electrons pass from the hot filament at the left between the plates a and b and strike a fluorescent screen at c. If there is a difference of potential between the plates a and b, the electron beam will be deflected from c.

by various types of suitably evacuated discharge tubes. That they really do carry negative charges can be shown by two different kinds of experiment. In one, first performed by Perrin (Fig. 120), the cathode rays are shot into a box-shaped electrode where they are trapped. This electrode is insulated and connected to an electroscope. The electroscope is found to charge up gradually; furthermore, it is found to be negatively charged. In another type of experiment which has many uses, a beam of cathode rays is shot between two plates a and b as shown in Fig. 121. If the plates are uncharged, the cathode-ray beam is

unaffected by their presence. But if the upper plate is positive and the lower negative, there is an electric field between them which will tend to drive positive charges downward and negative charges upward. By observing the change in the position c of the fluorescent spot where the cathode rays hit the glass or fluorescent material at the end of the tube, it can be proved that the cathode rays are negatively charged. This experiment can be done in lecture by using a cathode-ray oscillograph that has two pairs of plates, one vertical and one horizontal, so that the rays can be deflected up or down or to the right or left, depending on the plates used and their polarity. An analogous experiment can also be done with a charged water jet shot between two plates that are connected to the terminals of a Wimshurst machine.

Action of One Current on Another

5. The cathode-ray beam we have been discussing consists of charged particles moving rapidly along a clearly defined path. It is, therefore, an electric current. Since the charges are negative and moving away from the cathode, this current is equivalent to a positive current flowing into the cathode and of magnitude

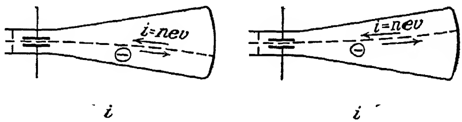


FIG. 122.—Deflection of a cathode-ray beam by a current in a wire parallel to it. The current $i = nev$ is, by convention, opposite in direction to the flow of electrons.

nev (Chap. XIII, Par. 17), where e is the electronic charge, v the velocity of the electrons, and n the number of electrons per unit length of the beam. Not only is such a cathode-ray beam an electric current, but it is one that is not confined to a heavy conductor. It responds easily to small forces, and any distortion of its path can be easily observed. Therefore it will be useful as a tool for studying small or rapidly changing electric forces. We have used it to show the deflecting effect of neighboring charges at rest on condenser plates. Let us next use it to show the effect of neighboring charges in motion, *i.e.*, of a neighboring current. Probably the ideal way of doing this experiment would be to use a second beam of cathode rays as the neighboring current, but

this is impracticable. For the second current, we use instead a straight copper wire parallel to the original direction of the cathode-ray beam and carrying a current of several amperes. As is shown in Fig. 122, if the current is running in the wire in the same direction as the current carried by the cathode rays (*i.e.*, in the opposite direction to the motion of the cathode rays themselves), the cathode-ray beam is pulled toward the wire. If the two currents are in opposite directions, the cathode-ray

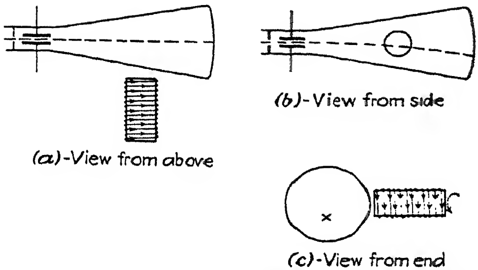
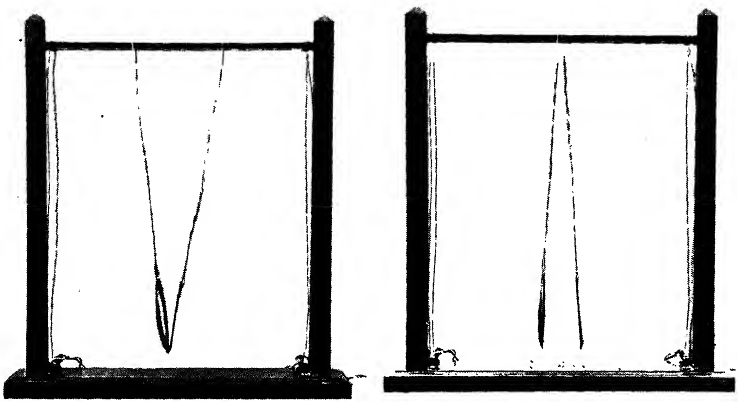


FIG. 123.—Deflection of a cathode-ray beam by the current in a solenoid. The direction of deflection is mutually perpendicular to the axis of the solenoid and the direction of the cathode-ray beam.

beam is repelled by the wire. Now substitute a helical coil for the straight wire. Such a coil is called a solenoid. If it is brought up in a horizontal plane with its axis perpendicular to the cathode rays, these are deflected up or down in the vertical plane, according to the direction in which the current is flowing in the solenoid. Schematic diagrams of this effect are shown in Fig. 123.

6. All these experiments show that currents in wires produce forces acting on a neighboring cathode-ray beam. According to Newton's third law, to every action there is an equal and opposite reaction so that the cathode-ray beam must also exert a force on the current-carrying wires. The inertia of the wires is so much larger than that of the cathode-ray beam that no motion is noticeable even though the force is the same. To observe the action and reaction, we shall use two wire-borne currents. For example, if we use straight parallel wires, we get effects exactly like those of Fig. 122 except that both wires move. As another example, suppose two flat coils of wire are suspended so

that they hang parallel to each other in vertical planes (Fig. 124). Then connect them to the terminals of a large battery so that large currents, 20 amp. or so, can be sent through them and put in a reversing switch so that the currents can be made to run in the same direction in both coils or in opposite directions. When the currents are in the same direction, the coils are seen to attract



(a) (b)

FIG. 124.—Two flat circular coils carrying currents. (a) Attraction when the currents are in the same direction. (b) Repulsion when the currents are in opposite directions.

each other. When the currents are in opposite directions, the coils repel each other.

7. Similar effects can be observed with two solenoids, one suspended so that it can turn in a horizontal plane, the other on a handle with flexible connecting wires so that it can be moved around at will (Fig. 125). To be specific, if we call the ends of the suspended solenoid A and B and of the other solenoid A' and B', it is observed that for one direction of current through the solenoids A repels A' but attracts B' while B attracts A' and repels B'. If the current through either solenoid is reversed, these effects are exactly reversed. If the currents through both solenoids are reversed simultaneously, the repulsions and attractions are unchanged.

8. We have described all these experiments in a qualitative way. There is no reason why they should not be made quantitative. We have already defined a unit of current and, of course, units of force. We could make a quantitative determination of the force between two wires and find out how it depended on the distance between the wires and the currents flowing in them. We would find that the force depended on the current flowing in each wire, on the length and relative orientation of the wires,

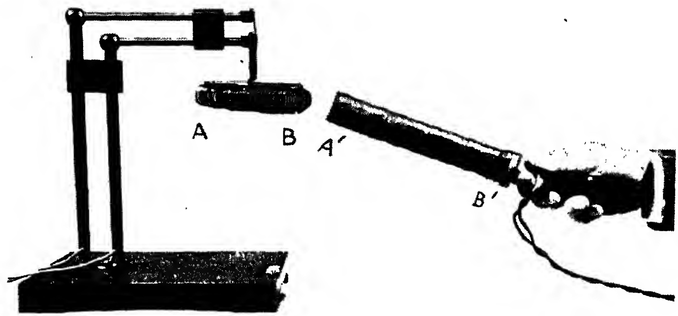


FIG. 125.—Reactions between two solenoids carrying currents.

and, inversely, on the distance between them. We might deduce a general law that would be applicable to all shapes and combinations of wires. Actually, we do not do so. Instead, we introduce an additional group of ideas, the ideas of magnets and magnetic fields and the interactions between magnets and electric currents. We do this for two reasons. In the first place, it makes the subject easier. The idea of a magnetic field is of great assistance both qualitatively and quantitatively. In the second place, there occur in nature many electric currents which do not manifest themselves directly as such, but are made evident by their magnetic effects. For example, the remarkable properties with which natural and artificial magnets are endowed have been used and studied for many centuries, although their interpretation in terms of the tiny currents caused by the motions of electrons in the atoms of which the magnets are composed is a recent affair and may not be entirely adequate.

9. Assuming that you all know, in a general way, what a magnet is, we shall first show that a simple bar magnet produces the same sort of effects on a current that are produced by another current. We can repeat the experiments of Pars. 5 and 6, showing that a magnet deflects cathode rays, that it repels or attracts a flat coil carrying a current, and so on. It behaves, in fact, exactly like a solenoid of similar size and shape with a very large current flowing through it. Either one will deflect a beam of cathode rays. One end of either one will attract one end and repel the other end of another solenoid or of another magnet. Once we have demonstrated this equivalence, let us turn to a study of magnets by themselves in order to establish some ideas that we can then apply to the study of currents.

SUMMARY

If a high potential difference is applied to two electrodes in an almost completely evacuated glass tube, a luminous discharge passes between them. If the pressure is low enough, the most conspicuous feature of this discharge is a beam of rays emerging in straight lines from the negative electrode. These rays are called cathode rays and are found to be rapidly moving light negatively charged particles. A beam of cathode rays is therefore a current but one of so small inertia that it responds to small forces.

Experiments with cathode rays and wire-borne currents show that parallel currents attract each other if they are in the same direction, repel each other if in opposite directions.

A cathode-ray beam is affected by a magnet in exactly the way it is affected by a current in a solenoid (helical coil). Solenoids and bar magnets are equivalent. In general, the forces between currents can be studied best in terms of magnetic effects.

ILLUSTRATIVE PROBLEM

What force acts on an electron passing between two plates of a cathode-ray oscillograph 5 mm. apart with a difference of potential of 100 volts between them? What acceleration does this give the electron? How long would it take to travel the distance between the plates?

We saw in Chap. XI that the electric field strength is the force per unit charge placed in the field. The force on charge e is therefore

$$F = Ee$$

where E is the strength of the electric field. The difference of potential is the work per unit charge done in carrying a positive charge from one point to the other. This is equal to the average force on unit charge times the distance. We have seen on page 231 that the field between the plates of a condenser is uniform so that the force is constant and

$$V = Ed.$$

The force per unit charge is therefore $E = V/d$, and the force on charge e

$$F = \frac{V}{d}e$$

$$F = \frac{100 \text{ volts}}{5 \times 10^{-3} \text{ m.}} 1.60 \times 10^{-19} \text{ coulomb}$$

$$= 3.20 \times 10^{-15} \frac{\text{joule}}{\text{m.}} = 3.20 \times 10^{-15} \text{ newton.}$$

The acceleration given by this force to the electron is found from $F = ma$ or

$$a = \frac{F}{m} = \frac{3.20 \times 10^{-15}}{9 \times 10^{-31} \text{ kg.}} \text{ newton}$$

where 9×10^{-31} kg. is the mass of the electron (see Appendix, Table 22).

$$a = 3.56 \times 10^{15} \text{ m./sec.}^2$$

The time required to go a distance of 5 mm. from rest with this acceleration is given by

$$s = \tfrac{1}{2}at^2$$

$$5 \times 10^{-3} \text{ m.} = \frac{3.56}{2} \times 10^{15} \text{ m./sec.}^2 \times t^2$$

$$t = \sqrt{\frac{10^{-17}}{3.56} \text{ sec.}^2} = 1.676 \times 10^{-9} \text{ sec.}$$

PROBLEMS

1. The vacuum in a discharge tube is sufficient so that ions and electrons do not collide in passing down the tube. The difference of potential between the electrodes is 10,000 volts. What velocity will an oxygen atomic ion have due to the electric field after traveling the length of the tube? An electron?

2. If the discharge tube of Prob. 1 is 25 cm. long, what force acts on the oxygen ion? What is its acceleration?

3. If the air in the discharge tube is under normal conditions of temperature and pressure, on the average, what velocity will an O_2 molecular ion gain from the electric field between collisions? What energy does this give the ion to knock electrons off molecules?

4. How many singly charged ions pass a point in 1 sec. in a current of 50 milliamp.?

5. If a proton is traveling with a velocity of 2×10^7 m./sec., through what potential difference must it have fallen?

6. An electron is accelerated in a cathode-ray-oscillograph tube by a potential difference of 300 volts. If the length of the deflecting plates in the direction of the beam is 1 cm. and they are 5 mm. apart with a potential difference of 100 volts between them, how far will the electron be deflected perpendicular to its path in passing between the plates?

7. If the electrons have the same velocity as they did in Prob. 6 and give a current of 1 microamp., how many electrons are there in unit length of the beam?

CHAPTER XV

MAGNETS AND MAGNETIC FIELDS

1. In the last chapter we showed that one electric current acts on another with a force that depends on the directions of flow of the currents, their magnitudes, and the distance between them. This is a new force, different from the electrostatic force between the charges carrying the currents. It is very much larger than this electrostatic force for currents of the ordinary type where the number of positive and negative ions in any section of the conductor carrying the current is approximately the same. We said that we called these forces that appear when charges begin to move, magnetic forces. We saw that the particular examples of this type of force shown by the reactions of two current-carrying solenoids on each other could be reproduced exactly by substituting bar magnets for one or both of the solenoids, and we suggested that the study of magnets and their reactions on one another might help us in studying the forces between electric currents. Magnets and magnetism form the principal subject of the present chapter but before taking them up directly it may be well to consider the equivalence between solenoids and magnets more fully. It was clearly demonstrated to be an experimental fact. Is it accidental, or can we show that it is to be expected from other phenomena and theories that we have already discussed?

2. The cathode rays described at the beginning of the last chapter are found to be independent of the nature of the gas originally in the discharge tube. We shall see in Chap. XIX that they are identical with the electrons whose existence we have postulated from time to time and can be obtained from every element. Thus there are electrons present in all forms of matter. We have become familiar with the ideas of the kinetic theory according to which the molecules of which matter is composed are in constant motion. It is not unreasonable, then, to assume that some or all of the electrons in atoms are themselves in

motion. We may imagine that these electrons are moving in closed orbits about the center of an atom very much as the planets move in closed orbits about the sun. Each electron moving in such an orbit is equivalent to an electric current moving in a tiny loop of perfectly conducting wire. Or an electron may be spinning on its axis like the earth; again, it is in motion so that it is equivalent to a small circular current. Perhaps the electrons are executing both motions simultaneously as the earth does. These moving electrons will exert magnetic forces on the other moving electrons in the same atom and in neighboring atoms and will respond to the magnetic forces of these other electrons or to the magnetic forces set up by electric currents flowing in neighboring conductors. In some atoms, the various moving electrons may be oriented in such a way that they compensate each other's effects so that the magnetic forces exerted by the atom as a whole are negligible. In other kinds of atoms, the effects of the different electrons do not compensate and the interatomic forces are so strong that they affect the orientations of the moving electrons in adjacent atoms either individually or by affecting the atom as a whole. We can think of each moving electron or system of electrons as equivalent to a small solenoid, and we can get some notion of what behavior to expect in a group of molecules by considering a model made of a number of such solenoids mounted in such a way as to be free to rotate about a diameter of the central turn.

3. It may be remarked parenthetically that we obviously would not go to the trouble of setting up this elaborate picture of the magnetic behavior of the molecules of which matter is composed if it were not going to be useful. It does explain the magnetic behavior observed in the large assemblages of molecules that can actually be studied. This is another instance of putting the cart before the horse for instructional purposes. Historically, the principal natural phenomena of magnetism were described pretty accurately by Gilbert, physician to Queen Elizabeth, but the theory suggested above was first foreshadowed by Ampère about a hundred years ago and has been worked out at all completely only in the last few years.

4. Returning to our model, suppose that we have enough small solenoids to represent crudely a bar of iron or other magnetic material and that initially the solenoids are oriented at random.

With no current flowing, the solenoids are equivalent to molecules that exert no magnetic force. Then, if a small current is turned on, the solenoids become equivalent to molecules with uncompensated electronic currents. In spite of the considerable

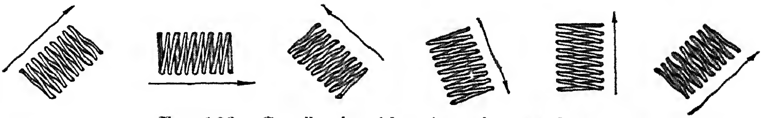


FIG. 126.—Small solenoids oriented at random.

magnetic forces now present, the friction and inertia of the solenoids prevent them from turning. Their resultant magnetic effect at any considerable distance away is negligible. If now the current is turned on in some large solenoid near them, its

FIG. 127.—Small solenoids aligned.

nearer end will attract one end of each of the little solenoids. Consider one row of these. Originally they are oriented at random as shown in Fig. 126, but if the large solenoid is at one end of the row, the current in it will produce an alignment of the

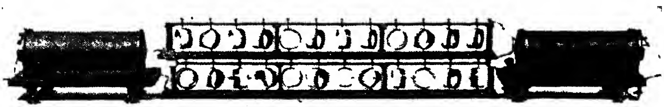


FIG. 128.—Photograph of two rows of small pivoted coils oriented at random. When the current flows in the large solenoids at each end of the rows, the coils are aligned by the magnetic effect of the large solenoids.

small solenoids like that in Fig. 127. Once this alignment has been produced by the action of the large external current, the little solenoids themselves will tend to preserve it since there will be forces of attraction between their adjacent ends. Since they are now all pointing in the same direction, their external magnetic effects will reinforce each other and the resultant will no longer be negligible. They will, in fact, be approximately equivalent to a single long thin solenoid.

5. If our model is to be even approximately like the molecules in a bar of magnetic material, it must contain not only one row

of pivoted coils but several. We can show that they will still be the equivalent of a single solenoid once they have been lined up. Suppose we have a number of rows like that in Fig. 127. Each is like a single solenoid, and therefore we have the equivalent of several long solenoids laid side by side. Then in cross section they look like Fig. 129, where the arrows indicate the direction of the flow of current. This set of circular currents will be essentially the same as the set of square currents in Fig. 130. But if the same current is flowing in each coil and we think of the coils as being packed close to each other, then it is evident

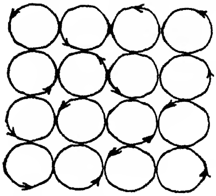


FIG. 129.—A cross section of several rows of small aligned solenoids showing the direction of the currents in each row.

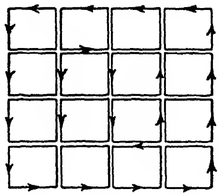


FIG. 130.—A cross section of current squares approximately equivalent to the circular currents of Fig. 129. The magnetic effect of the currents in adjacent wires cancels.

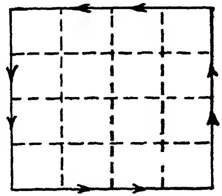


FIG. 131.—Resultant effective current of a set of solenoids like that shown in Fig. 130. The small solenoids are equivalent to a single large one.

that everywhere, except on the outside, each little piece of current-carrying wire is immediately adjacent to another piece of wire carrying an equal current in the opposite direction. The resultant effect is as if no current were flowing except on the outside of the set of solenoids as shown in Fig. 131. Thus we see that a set of long solenoids of small cross section laid beside each other is equivalent to a single solenoid of the same length and number of turns and of a cross section equal to the sum of the cross sections of the component solenoids. Therefore, we have shown that many small solenoids originally pointing in all directions become equivalent to a single large solenoid when they are aligned by the magnetic influence of an external current.

6. Now turn to a consideration of the magnetic material of which the set of turning solenoids is supposed to be a crude model. Take a bar of iron for example. The atoms in the bar are vibrating and turning around positions of equilibrium under the influence of the thermal agitation. This chaotic motion will

always resist a force that tends to keep the atoms or the electrons in them in a definite alignment, but if such a force is sufficiently strong such an alignment may be brought about. As soon as such an alignment has been brought about, the magnetic effects of all the atomic currents in the bar will be in the same direction so that their cumulative effect may, and in iron usually will, be strong enough to be felt outside the bar. In short, the bar will have become magnetized or, as we say, become a magnet. In other words, it will now behave very much like a single solenoid of similar dimensions just as the assemblage of solenoids in our model behaved like a single solenoid after alignment had occurred. Furthermore, the mutual interaction of the atomic solenoids will tend to keep them in alignment in spite of the disturbing effect of thermal motion. The degree to which the alignment persists, *i.e.*, the degree of permanence of the magnetization, is found to depend very much on the purity and metallurgical treatment of the iron. It is possible to make permanent magnets that will remain strongly magnetic indefinitely. Such magnets are always made of iron or some alloy of iron since no other common element can be nearly so effectively magnetized. Since a permanent magnet acts in exactly the same way as a current-carrying solenoid, it can itself be used to magnetize other pieces of iron. Evidently the question posed in the first paragraph of this chapter has been answered. The equivalence between magnets and solenoids is understandable in terms of the atomic and electrical structure of matter.

7. Not only iron itself but some of its compounds have the property of being magnetizable. It is in specimens of minerals containing these compounds that we find a state of magnetization existing in nature. The substance most commonly found in this condition is magnetite (Fe_3O_4) which is one of the important ores for the commercial production of iron. A piece of magnetite that is magnetic is called a lodestone. By using electric currents, it is now possible to produce permanent magnets so much better than lodestone or iron magnetized by lodestone that many a modern physicist would not recognize a lodestone if he saw one.

The Effect of Temperature

8. If the molecular agitation is increased, it is evident that the permanent alignment of molecular solenoids will become more

difficult. We might expect to destroy the magnetization of a piece of iron by heating it to a sufficiently high temperature. This expectation can be easily verified experimentally. The temperature at which this effect begins to predominate is called the Curie point and is about 500°C. for iron. For some alloys it is much lower, in some cases a little above room temperature.

Magnetic Poles and Magnetic Fields

9. Granted that we have shown that magnetic effects can be interpreted in terms of electric currents, let us forget about this interpretation for the moment and put ourselves in the position of the early physicists who could work more easily with magnets than with electric currents. Suppose that we have plenty of permanent magnets of various shapes and sizes and also plenty of unmagnetized iron and iron filings available but that our knowledge of electricity is limited to static electricity. Let us repeat the experiments described on page 301, Chap. XIV, using permanent magnets instead of solenoids. We saw that these experiments were very similar to those on static electricity which led to Coulomb's inverse-square law of force between electric charges. We find that this resemblance is more than superficial. We find that the interactions between different magnets can be explained if we assume that there are centers of force near the two ends of a bar magnet. These centers of force are called magnetic poles and are of two kinds which we call north and south. In any given magnet, if there is a north pole of a certain strength at one end, there is a south pole of equal strength at the other end. The names north and south are derived from the fact that the earth is itself a magnet and will cause any magnet that is free to turn to align itself roughly in a north-south direction. The end that points south is said to contain a south pole. For purposes of calculation, it is convenient to call north poles positive and south poles negative. The law of force that is found to hold between magnetic poles is again the inverse-square law which we have already encountered in the realms of gravitation and electricity. Stated explicitly, it is as follows:

The force between two magnetic poles is proportional to the product of their pole strengths and inversely proportional to the square of the distance between them; it is an attraction if they are unlike poles and a repulsion if they are like poles.

Stated mathematically, it is

$$\frac{pp'}{d^2} \tag{1}$$

where F is the force, p and p' are the pole strengths, and d is the distance between them. F is a force of repulsion if it is positive, of attraction if it is negative. In Par. 19, we shall discuss this law again in terms of an exact equation. For the moment, we leave it as a proportion.

10. Since the law of force between magnetic poles is of the same form as that between electric charges and as that between

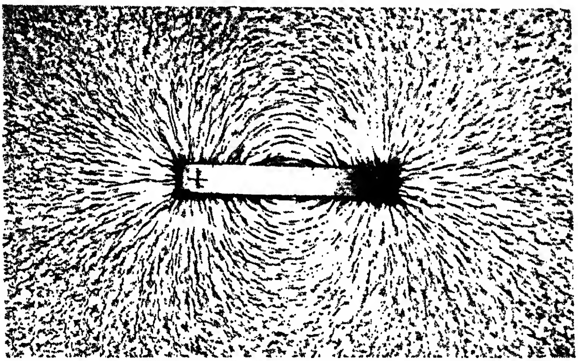


FIG. 132.—The magnetic field of a single bar magnet.

gravitational masses, it is evident that many of the considerations of Chap. XI can be extended to the problems of magnetism. A region where the effects of neighboring magnetic poles are felt can be defined as a field of magnetic force, and the strength of the magnetic field at various points in the region will express the relative forces that would act on a given pole placed at these points. Like the electric field strength, the magnetic field strength is evidently a vector quantity.

11. As in the case of electric fields, we can learn about a field of magnetic force without knowing the location or magnitude of the magnetic poles that give rise to it. This is hardly surprising when we remember that we originally introduced the notion of such poles as a convenience to help describe the reactions between electric currents. We do not pretend that a magnetic pole has

the same degree of ultimate reality that an electric charge has. There are, in fact, very real differences in their behavior which are understandable immediately in terms of the origin of magnetic poles. In talking about electric charges, we frequently spoke of an isolated body as carrying a positive or negative charge and of charges as flowing through conductors from one point to another in space. There is no difficulty about separating the two kinds of charges or about making negative charges flow through solid conductors. In magnetized bodies, on the other hand, we find that

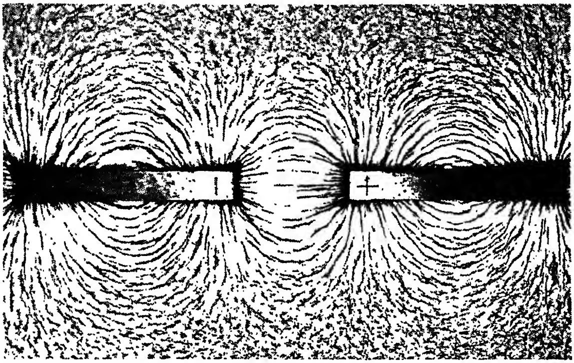


Fig. 133.—The magnetic field of two bar magnets with unlike poles facing each other.

whenever there is a positive pole present there is always a negative pole of equal strength somewhere else in the body. For example, a simple bar magnet has a positive pole at one end and a negative pole of equal strength at the other end. If it is cut in two, equal and opposite poles appear at the severed faces so that we have two bar magnets each half as long as the original but with the same pole strengths. This process can be continued indefinitely without ever isolating a pole of one kind. That this must be so is immediately obvious if we consider the equivalent solenoid. The two kinds of polarity merely represent the two ends of the solenoid, and dividing the solenoids will result in the appearance of two more ends or poles. We cannot have a solenoid with only one end and therefore cannot have only one kind of pole. We might bend a long solenoid into a ring but we would then have no free pole at all. The moment such a ring was cut, a pole of

each sign would appear immediately. Similar arguments can be followed through for magnetized bodies of any shape.

Magnetic Lines of Force

12. It is often convenient to have a general qualitative notion of the distribution of the magnetic field throughout the region surrounding a system of poles even though the exact value at every point is not calculated. For this purpose, we make use of the idea of lines of force which has already been introduced in

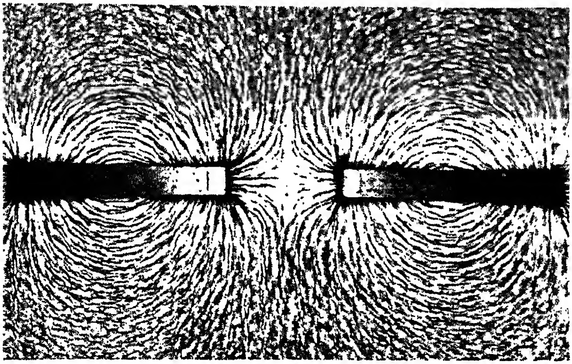


FIG. 134.—The magnetic field of two bar magnets with like poles facing each other.

the study of electric fields. In magnetic fields as in electric fields, a line of force is a line that is at every point tangent to the direction of the field strength at that point. Lines of magnetic force appear to come out of north poles and end on south poles. Furthermore, if we make the rule that no lines of force start or stop except where there are poles and that the same number of lines always starts from any unit pole, then the number of lines of force in any region is proportional to the magnetic field strength in that region. Since we have seen that it is impossible to have one magnetic pole without an equal and opposite pole on the same body, it is evident that the number of lines of magnetic force starting and ending on any one body must be the same. This suggests that these lines of force do not end on the surface of the body but simply disappear into it. That they are in fact always closed lines encircling electric current will appear as soon as we

study the magnetic fields set up by electric currents. This is evidently consistent with the view of Par. 11 that magnetic poles are convenient fictions. They might almost be defined as

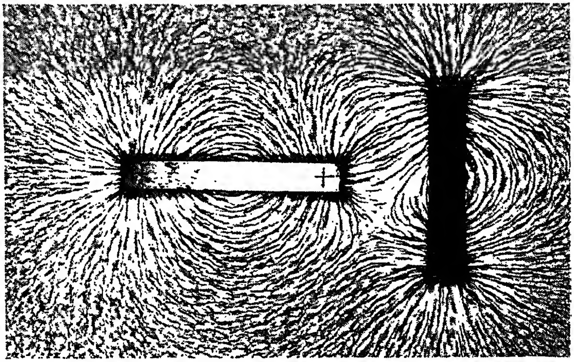


Fig. 135.—The magnetic field of two bar magnets at right angles to each other.

the places where many lines of magnetic field strength disappear into magnetized bodies or emerge from them.

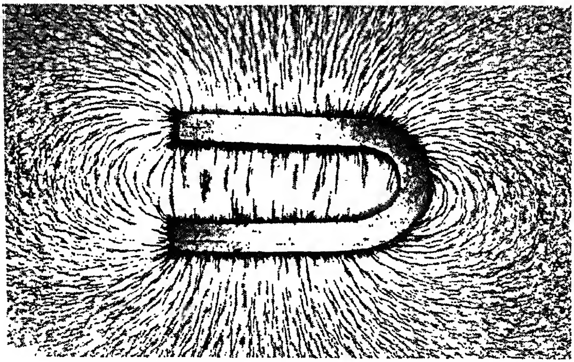


Fig. 136.—The magnetic field of a horseshoe magnet.

13. The process of finding out the direction and number of the lines of magnetic force in a region is called mapping the magnetic field in that region. Exact results require elaborate calculations, but in many cases it is possible to get fairly accurate ideas by using iron filings to indicate the nature of the field. Each filing

tends to become magnetized by the field and, therefore, to line up parallel to it. Also the little magnets so formed attract each other. Consequently, the filings tend to form themselves in lines corresponding rather closely to the true lines of force. Examples of this effect are shown in Figs. 132 to 136.

Magnetic Fields from Electric Currents. Unit Magnetic Field

14. In the last chapter, we described the forces of interaction between currents without using the concepts of magnetism. In the present chapter, we first discussed the possible existence of electric currents within the atoms of which matter is composed and the effects that such currents might cause. Then we pointed out that such effects were familiar in the realm of magnetism. Subsequently, in dealing with the phenomena of magnetism, we have deliberately avoided any discussion of their electric origin. In order to make the concepts of magnetism as clear as possible, we have spoken of magnetic poles as if they had a real existence quite apart from any electrical effects and of magnetic fields as arising from the presence of magnetic poles. The magnetic fields from natural or artificial magnets have their ultimate origin in electric currents that cannot be directly studied or observed. Therefore it is convenient to introduce the concept of magnetic poles and speak of magnetic lines of force as beginning and ending on these poles. But we have seen that a beam of free electrons (cathode rays) or the flow of electrons in a wire such as constitutes an ordinary current also produces magnetic effects. These magnetic effects are interpreted in terms of the magnetic fields produced by the current. As we shall see, it is not necessary to use the concept of a magnetic pole at all, and in many cases, it would be difficult to find the distribution of poles that would give rise to the field set up by a particular distribution of electric currents. But the concept of a magnetic field of force is a very general one and a very useful one. The whole theory of electrical machinery is based on the interplay between electric currents and magnetic fields.

15. So far we have considered in a qualitative way the magnetic effects of solenoids and of straight wires in terms of their action on other currents. We can study the magnetic fields produced by such currents by mapping the field with iron filings such as we used for studying magnets. The fields produced by a

single straight wire and by a solenoid are shown in Figs. 137 and 138. We see that the lines of magnetic force are closed lines

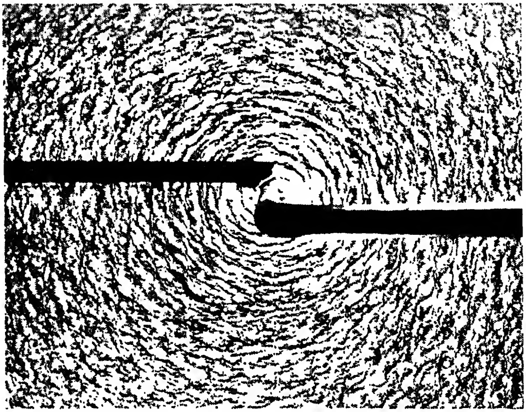


FIG. 137.—The magnetic field around a straight wire. (The wire comes in from the right above the plane of the filings, turns at right angles to pass through this plane, turns again below and goes off at the left of the picture.)

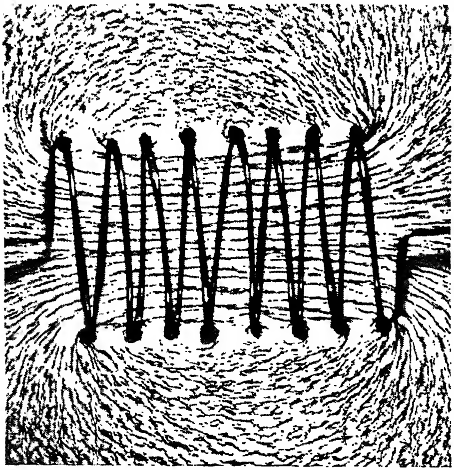


FIG. 138.—Magnetic field of a solenoid shown by iron filings in a plane through the axis of the solenoid.

encircling the current-carrying conductors as was anticipated in Par. 12.

16. Another form of current giving a very simple symmetrical type of magnetic field is that in a circular coil. The field from a one-turn coil of this type is shown in Fig. 139. Suppose we consider coils of this type of various dimensions and numbers of turns and with different currents running through them. A small magnet or compass needle can be used to test the magnetic field strength at the center of these different coils. The torque (see Par. 21 below) tending to turn the magnet into the direction

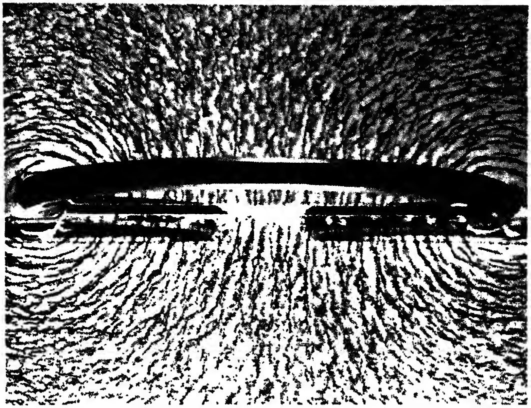


FIG. 139.—Magnetic field of a circular current. (The iron filings are on a transparent plate perpendicular to the plane of the circle and through its center. The parts of the circle above and below this plane show in the picture.)

perpendicular to the plane of the coil will be proportional to the force on its poles which is itself proportional to the magnetic field. To compare the effect of different coils and of different currents in the same coil, it is not necessary to know anything about the strength of the magnet but merely to use always the same magnet. The results of such experiments prove that the magnetic field at the center of a circular plane coil is proportional to the number of turns in the coil and to the current flowing in it and inversely proportional to the diameter of the coil. If the current is flowing counterclockwise, the direction of the field is out from the plane of the coil. Expressed mathematically, we have

$$H \propto \frac{nI}{D} \qquad \text{or} \qquad H = k\frac{nI}{D} \tag{2}$$

quantities in this expression that are characteristic of the particular magnet, and the torque depends only on their product. This product is called the moment of the magnet and can be designated by M. Substituting this in our previous expression, we find that the torque on a magnet in a uniform magnetic field is given by

$$T = HM \sin \theta \tag{6}$$

where H is the strength of the magnetic field, M the moment of the magnet, and θ the angle between them. This torque can be measured, and we can define a magnet of unit moment as one on which unit magnetic field exerts unit torque when the field and magnet are at right angles. This is the equivalent of our previous definition of the unit of magnetic pole strength.

Magnetic Potential

22. The definitions of potential difference and potential in Chap. XI depended on the idea of work done in moving a unit charge from one point to another in an electric field of force. It is clear that we can set up a similar definition of the magnetic difference of potential although the absence of magnetic conduction currents makes the idea of less general importance. A magnetic difference of potential between two points can be defined, then, as the work per unit pole required to move a positive pole from one of the points to the other. Since the definition of difference of potential is the same for magnetism as for electricity and the law of force is also the same, the results obtained in electrostatics can be carried over into magnetism.

Magnetic Dipoles

23. We saw in Chap. XI that it followed directly from Coulomb's inverse-square law that the electric field caused by a charged sphere at any point in the surrounding space was given by $Q/4\pi\epsilon r^2$ where Q was the charge on the sphere, r was the distance from the center of the sphere to the point in question, and ϵ was the dielectric constant. The corresponding expression for the potential was $Q/4\pi\epsilon r$. By a similar argument, we can show that the magnetic field caused by a pole is $p/4\pi\mu r^2$ where p is the strength of the pole, r the distance, and μ the permeability. Similarly, the magnetic potential is $p/4\pi\mu r$. But we have seen

that we are never concerned with an isolated pole; in practice, we are always dealing with pairs of poles. Such a combination of two equal and opposite poles is completely analogous to the electric doublet or dipole which we considered at length in Pars. 28, 29 and 30 of Chap. XI. We have already defined the moment of a magnet in Par. 21. If we substitute this for the electric moment and make similar appropriate substitutions of magnetic for electric quantities, the results of Chap. XI can be applied directly to obtain the magnetic potential and magnetic field set up by a magnetic dipole. The only difference is that in magnetism we have nothing but dipoles, whereas in electricity they constitute a rather unusual distribution of charge.

SUMMARY

The behavior of magnets can be interpreted in terms of electric currents caused by the motion of electrons in molecules. From this point of view, all magnetic effects are essentially electrical effects. The process of magnetization is an alignment of the atomic circuits in such a way that their magnetic effects reinforce each other. This is done most easily and completely in iron and some of its compounds. Alignment is opposed by the thermal motion of the atoms, and a magnet loses its magnetic properties if its temperature is raised sufficiently.

Disregarding the electrical origin of magnetism, the observed effects can be ascribed to centers of force called magnetic poles. These are of two kinds, positive and negative, and always occur in pairs. The force between magnetic poles obeys a law exactly analogous to that for the force between two charges. Consequently, there are magnetic fields of force and magnetic lines of force. A magnetic field can be mapped by the use of iron filings.

The magnetic field of an electric current flowing in a plane circular coil is found to be proportional to the current and the number of turns in the coil and inversely to the diameter. Unit magnetic field is then defined as the field at the center of a plane coil of one turn with a diameter of one meter when a current of one ampere is flowing through it. Unit magnetic pole is then one on which a unit field will act with a force of one newton.

The moment of a bar magnet is the product of its pole strength and the distance between its poles. The torque on a bar magnet of moment M in a field H is $HM \sin \theta$ where θ is the angle between

the direction of the field and the magnet. The magnetic potential and magnetic field strength from a short bar magnet or magnetic dipole can be calculated in the same way as for an electric dipole.

ILLUSTRATIVE PROBLEMS

1. Calculate the field strength at the center of a circular coil that consists of five turns 35 cm. in diameter and carries a current of 40 amp.

The field strength is given by Eq. (3), page 318, as

$$H = \frac{nI}{2a}$$

where $n = 5$ = the number of turns.
$I = 40$ amp. = the current in each turn.
$a = \frac{0.35}{2}$ m. = the radius of the circle.

$$H = \frac{5 \text{ turns} \times 40 \text{ amp.}}{2 \times 0.35/2 \text{ m.}} = \frac{571 \text{ amp. turns.}}{\text{m.}}$$

2. Two magnetic dipoles each have a moment of

$$2 \times 10^{-4} \frac{\text{newton-m.}^2}{\text{amp. turn}}.$$

One dipole is placed on the axis of the other with their centers 20 cm. apart. What angle must the second dipole make with respect to the axis of the first in order that it experience a torque of 0.2 newton-m.?

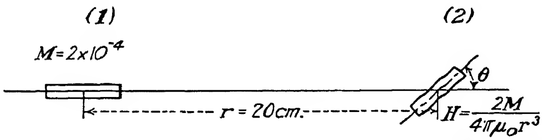


FIG. 141.—The orientation of the second dipole in the magnetic field of the first is given by the angle θ. The problem is to find θ so that the torque on the second dipole will be 0.2 newton-m.

The field in which the second dipole finds itself (Fig. 141) due to the first is given by analogy with Eq. (19), page 248, as

$$H = \frac{2M}{4\pi\mu_0 r^3}$$

where $M = 2 \times 10^{-4} \frac{\text{newton-m.}^2}{\text{amp. turn}}$ = the moment of the dipole producing the field.
$r = 0.20$ m. = the distance between centers of the dipoles.

$$\mu_0 = 4\pi \times 10^{-7}\frac{\text{newton}}{(\text{amp. turn})^2} = \text{the permeability of empty space.}$$

$$H = \frac{2 \times 2 \times 10^{-4}\dfrac{\text{newton-m.}^2}{\text{amp. turn}}}{4\pi \times 4\pi \times 10^{-7}\dfrac{\text{newton}}{(\text{amp. turn})^2}(0.20\ \text{m.})^2} = 3{,}170\frac{\text{amp. turn}}{\text{m.}}.$$

The torque on a magnet in a magnetic field is given by Eq. (6), page 320, as

$$T = HM \sin \theta$$

where $T = 0.2$ newton-m. = the torque.

$H = 3{,}170\dfrac{\text{amp. turn}}{\text{m.}}$ = the strength of the magnetic field.

$M = 2 \times 10^{-4}\dfrac{\text{newton-m.}^2}{\text{amp. turn}}$ = the moment of the magnet.

θ = the angle which the magnet makes with the magnetic field.

$$0.2\ \text{newton-m.} = 3{,}170\frac{\text{amp. turn}}{\text{m.}} \times 2 \times 10^{-4}\frac{\text{newton-m.}^2}{\text{amp. turn}} \sin \theta.$$

$$\sin \theta = \frac{10^3}{3{,}170} = 0.315.$$

$$\theta = \arcsin 0.315 = 18°22'.$$

PROBLEMS

1. What is the strength of the magnetic field at the center of a circular coil 10 cm. in diameter of 100 turns carrying a current of 5 amp.?

2. A wire 3 m. long is made into a circular coil with a mean radius of 6 cm. Find the strength of the field produced at the center of the coil by a current of 0.1 amp. in the wire.

3. What must be the diameter of a coil of three turns of wire in order that a current of 5 amp. may produce a magnetic field of 12 amp. turns/m. at its center?

4. Two 10-cm. magnets are laid with their north poles opposing and their centers 20 cm. apart on a straight line. One magnet has poles of strength 5×10^{-5} weber and the other of strength 10^{-5} weber. With what force do they repel each other?

5. A 50-g. magnet floats in the air 2 cm. above an equal magnet as in Mohammed's coffin. If the magnet is 10 cm. long, what must be the strength of one of its poles?

6. What is the force of compression on the steel of a 10-cm. magnet with pole strength of 5×10^{-5} weber? If the magnet is 1 sq. cm. in cross section, how much is it compressed by this force?

7. What is the moment of the couple acting on a magnet of moment 3×10^{-5} weber-m. placed at right angles to a magnetic field of 4 amp. turn/m.?

8. What torque acts on a dipole of moment 2×10^{-7} weber-m. due to the horizontal component of the earth's magnetic field of 13 amp. turns/m. if the dipole axis makes an angle of 45° with the horizontal component of the field?

9. A 5-cm. magnet has a pole strength of 6×10^{-5} weber. The torque on this magnet is 0.1 newton-m. when placed at an angle of 60° with a magnetic field. What is the strength of the field?

10. A small magnetic dipole of moment 3×10^{-6} weber-m. is placed at the center of a circular coil of 100 turns and radius 10 cm. How much current must there be in this coil to produce a torque of 5×10^{-3} newton-m. if the dipole makes an angle of 60° with the plane of the coil?

CHAPTER XVI

ELECTRIC CURRENTS AND MAGNETIC FIELDS

1. In Chap. XIV, we discussed the interaction of electric currents in terms of some simple observations. Then in Chap. XV, the phenomena of permanent magnetism were described and interpreted in terms of submicroscopic currents. The concepts of magnetic fields and magnetic poles were defined and methods of mapping magnetic fields demonstrated. The magnetic field of a particular type of electric circuit was used to define a unit of magnetic field strength, and pictures were shown of the fields set

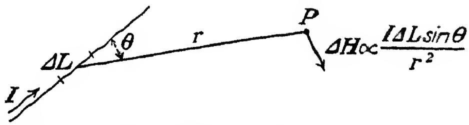


FIG. 142.—Ampère's law. The magnetic field at the point P due to the current element of length ΔL is given.

up by other kinds of circuits. In the present chapter, we shall discuss the general laws expressing the relations between electric currents and magnetic fields and show how they explain the kind of phenomena described in Chap. XIV. The first step in this process is the statement of the law for the magnetic field set up by any electric current. This is an experimental law, first formulated in general form by Ampère to explain his observations, and confirmed by the measurements of many subsequent experimenters. It is as follows:

Each element of length of a current-carrying conductor makes a contribution ΔH to the magnetic field at a point P, which is proportional to the length ΔL of the element, to the current I flowing in it, to the sine of the angle θ between the direction of the current and the line r connecting it with P and inversely proportional to the square of this length r. The magnetic field is perpendicular to the plane determined by ΔL and r; into this plane if θ measured from I to r is clockwise, out of the plane if θ is counterclockwise.

In the units we are using, the factor of proportion comes out $1/4\pi$ so that we have

$$\Delta H = \frac{I \Delta L \sin \theta}{4\pi r^2} \tag{1}$$

if I is in amperes, ΔL and r in meters, and ΔH in ampere turns per meter.

Magnetic Field of a Straight Wire

2. In Fig. 137 of the last chapter, we already have seen a roughly quantitative picture of the magnetic field around a straight wire. Let us use this picture as a test of Ampère's law. The iron filings in the photograph are on a sheet of mica perpen-

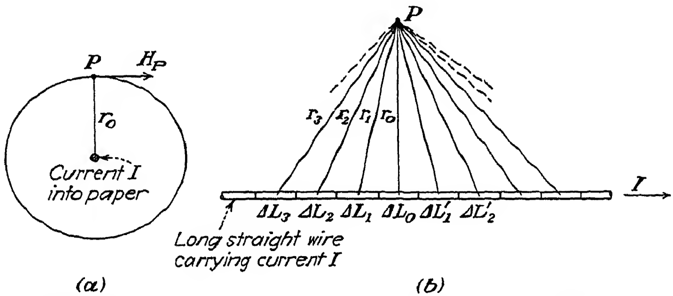


FIG. 143.—Application of Ampère's law to the case of a straight wire-borne current. (*a*) shows the magnetic field at the point P in a plane perpendicular to the wire. (*b*) shows the computation of the field at P from each short section ΔL of the wire.

dicular to the wire and are arranged in circles around the wire. Consider whether this is consistent with Ampère's law. A line drawn from the wire to any point P on the mica and the wire itself determine a plane in which lie the lines from P to every element ΔL of the wire. This plane is shown in Fig. 143(*b*). By Ampère's law, the contribution to the magnetic field at P from each element of the wire will be perpendicular to this plane, *i.e.*, in the same direction for the contribution of each element of the wire. The direction of the resultant field at P therefore is that shown in Fig. 143(*a*). Similarly for any other point in the plane of the mica, the resultant field will be in that plane and perpendicular to the line from the point to the wire. Thus if we take a series of points on a circle whose center is in the wire, the

field at each point will be tangent to the circle. Therefore the circle represents a line of magnetic force. Furthermore, without carrying out the calculation in detail, it is evident from Ampère's law that the field will be stronger close to the wire than far out from it since the average value of the r for the different parts of the wire will be smaller. Ampère's law, then, tells us that the magnetic lines of force should be circles and that they should be closer together near the wire than far away from it. These conclusions are in agreement with the arrangement of iron filings shown in Fig. 137 of the last chapter.

3. To get an exact expression for the magnetic field at any point resulting from the current in a straight wire, it is necessary to add up the contributions of all the little elements of length of the wire. From Fig. 143(b), it is evident that both the length of r and the angle it makes with the wire change for different parts of the wire. To take account of these changes and perform the summation of the effect of all parts of the wire requires the use of calculus. We shall merely quote the result, which is

$$H_p = \frac{I}{2\pi r} \tag{2}$$

for the magnetic field H_p, at the point P, at a perpendicular distance r, from an infinitely long straight wire carrying a current I. The direction of the field is best obtained by the help of the familiar "right-hand rule," *i.e.*, if the wire is grasped in the right hand with the thumb pointing along the wire in the direction of the current, the fingers will encircle the wire in the direction of the lines of magnetic force. As we have seen, the lines of force are circles in planes perpendicular to the wire.

Magnetic Field of a Circular Loop

4. We have already given the result for this case in our discussion of the unit of magnetic field strength. Nevertheless, it may be well to see whether the results we have quoted are consistent with the general law. Referring to Fig. 139 of the last chapter, we have a photograph of iron filings on a piece of mica perpendicular to a plane circular loop of wire and passing through a diameter of the circle. Their arrangement shows the magnetic field produced by a current flowing through the wire. Very close to the point where one side of the loop goes through the

mica, the effect of the current in that part of the wire predominates so that it acts almost like a long straight wire. The lines of force, therefore, are circles with their centers in the wire. Farther away, the effect of the other parts of the wire becomes important and the lines of force are closer together inside the loop than outside, have less curvature, and are no longer exactly circular. In the middle of the loop, they are perpendicular to its plane. To calculate the magnitude and direction of the resultant magnetic field at any point of this arrangement is difficult, but for the particular point at the center of the loop, conditions are so symmetrical that the calculation is easily carried out by elementary methods as follows.

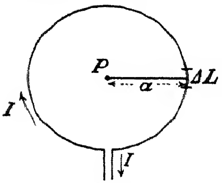


FIG. 144.—Application of Ampère's law to the current in a circular circuit.

5. Let the radius of the loop of wire be a. Now in applying Ampère's law, it is evident that the angle θ, the angle between the element of length ΔL of the wire and the line from that element to the point where the field is to be determined, is 90° for every part of the wire. Furthermore, the distance r is constant and equal to a for every part of the wire. Consequently, the contribution of each element of length ΔL to the field at the center is

$$\frac{I\Delta L}{4\pi a^2}$$

and these contributions have to be added for the whole circumference of the loop. This must be a vector addition, but since the contribution of every element in this particular case is in the same direction, perpendicular to the plane of the loop, vector addition is equivalent to ordinary arithmetical addition. Since a and I are both constant, we merely have to sum up ΔL around the loop and this summation gives us the whole circumference of the loop $2\pi a$. Therefore, the total magnetic field at the center is

$$H = \frac{I}{4\pi a^2} \times 2\pi a = \frac{I}{2a} \tag{3}$$

and is directed into the plane of the loop if the current is clockwise, out of the plane of the loop if the current is counterclockwise as is shown by applying the right-hand rule given in Par. 3.

Magnetic Field of a Coil of Several Turns

6. If instead of a single loop of wire, we have a coil of several turns, the resultant effect at the center of the coil will simply be the sum of the effects of the separate turns; and if their dimensions are all approximately the same, the field at the center will be

$$H = \frac{nI}{2a}$$

or

$$H = \frac{nI}{D} \tag{4}$$

where n is the number of turns, a their average radius, and D their diameter. This is in agreement with the experimental result given in the last chapter, Par. 17, Eq. (3).

Work Required to Move Unit Pole around a Current

7. We wish now to calculate the work that is done in moving a unit positive magnetic pole once around a current-carrying conductor. This will make it possible to calculate the magnetic field in a solenoid by elementary means. We dislike using so artificial a concept as an isolated pole, but we believe that for the beginning student it is preferable to the other possibilities. Return to the case of the infinitely long straight wire discussed in Par. 3, and suppose that we start with a unit positive pole at a distance r from the wire. Then suppose we move that pole in a circle whose center is in the wire and whose plane is perpendicular to the wire as shown in Fig. 145. The force at every point is H newtons by definition, and this, by Eq. (2), is $I/2\pi r$; and the distance through which the pole is moved to return it to its starting point is simply the circumference of the circle $2\pi r$. Therefore, if the pole is moved in the direction against the field, an amount of work equal to

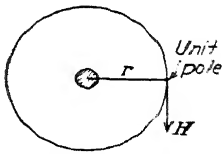


FIG. 145.—The work done in moving a unit pole around a current is the product of the force H and the distance $2\pi r$.

$$2\pi r \times \frac{I}{2\pi r} = I \text{ joules} \tag{5}$$

must be done on it to bring it completely around the wire. It is evident that this result is independent of the radius of the circle and, therefore, holds for any circular path. Any other path,

even if it is not all in one plane, can be shown to be equivalent to a path made of small segments of circles whose centers are in the wire, combined with elements of path parallel to the radii of these circles or parallel to the wire. Both these latter elements of path will be in directions perpendicular to the magnetic field and therefore against no force and therefore requiring no work. Consequently Eq. (5) is a perfectly general one and gives the amount of work required to move a unit positive pole around a straight current I by any closed path whatever. Furthermore, it applies to a current in a conductor of

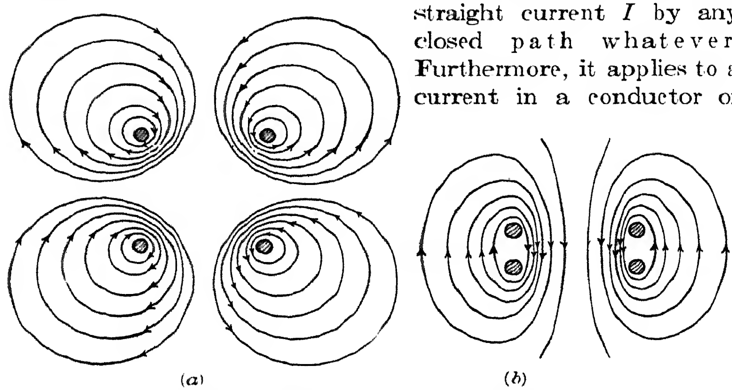


FIG. 146(*a*).—The magnetic field from two loops of wire some distance apart. The fields coincide along vertical axes inside and outside the coils and neutralize each other along a horizontal axis through the center.
FIG. 146(*b*).—The magnetic field from two loops of wire close together. The horizontal fields of Fig. 146(*a*) have disappeared.

any shape since the encircling path can always be taken so close to the wire that the field can be taken as given by Eq. (2).

The Magnetic Field of a Solenoid

8. We have given the discussion of the last paragraph largely to be able to calculate the magnetic field in a solenoid. The effect of a solenoid can be understood if we think of two flat coils with their planes parallel, at first some distance apart and then brought closer together. As is shown diagrammatically in Fig. 146(*a*), the effects of the two coils are in opposite directions in the region between the wires and, therefore, cancel each other, whereas along the axis and on the outside they are in the same direction and reinforce each other as is shown in Fig. 146(*b*). A solenoid is the equivalent of a whole series of coils piled on top of each

other and gives a resultant field of magnetic force such as is shown in Fig. 147 diagrammatically, and by iron filings in Fig. 138 of the last chapter. Outside the solenoid, this field is seen to be of exactly the same type as that given by a bar magnet; and, therefore, it is now understandable why the solenoid and the bar magnet gave identical results in the experiments described in Chap. XIV. In fact, the experiments there described can now

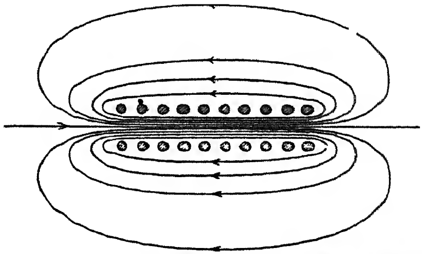


FIG. 147.—The magnetic field of a solenoid.

be interpreted in terms of magnetic fields. The student is urged to try this.

9. In order to get the quantitative expression for the magnetic field produced by a solenoid, we shall consider the special case of a closed ring solenoid such as is shown in Fig. 148. It is evident from the arrangement of the iron filings in Fig. 148 that there is no magnetic field on the outside of the ring. In terms of equivalent magnetic poles, the north and south ends of the equivalent bar magnet have been bent around until they are in contact and completely counteract each other. To get the magnetic field inside the solenoid, consider the work done in moving a unit pole around a circle inside the solenoid. This circle encloses all the wires on the inner side of the solenoid, and none of those on the outer side. The currents in the wires on the inner side are all in the same direction (into the paper if the magnetic field is clockwise). Therefore, the work done in moving the positive unit pole around the circle is NI joules if I is the current measured in amperes and N is the number of turns in the solenoid. But the distance that the pole has been moved is $2\pi R$, where R is the radius of the ring in which the solenoid is bent. If R is very large compared with the radius r of the solenoid winding, the field is approximately uniform inside the solenoid. If we call its

value H, the work done on the pole must have been $H \times 2\pi R$ and, therefore,

$$H = \frac{NI}{2\pi R}.$$

If we now replace N by n, the number of turns per unit length, which is equal to $N/2\pi R$, we get

$$H = nI \tag{6}$$

as the value of the magnetic field inside a solenoid of n turns per meter carrying I amp. of current.

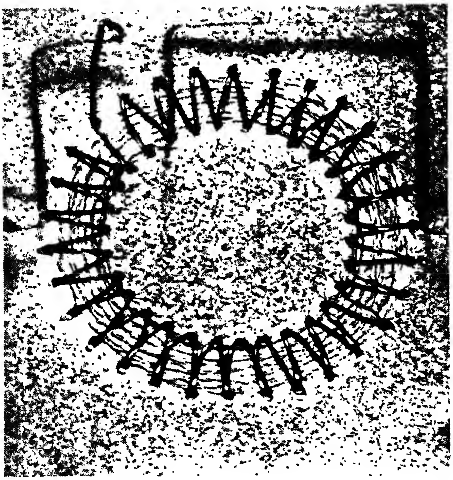


FIG. 148.—The magnetic field of a ring solenoid.

Equivalence between Moving Charges and Currents. Rowland's Experiment

10. It is to be hoped that the way in which electric currents were introduced in this book has made the student feel that they are obviously equivalent to charges in motion. But in the historical development of the subject this was not so. One reason that various systems of units arose was because of the seeming lack of connection between static electricity and current electricity. That charges in motion have the same magnetic effect as ordinary currents was first proved by Rowland at

Johns Hopkins in 1878. He mounted some pieces of tinfoil on the outer edge of a large insulating disk. These pieces of tinfoil could be charged and the disk revolved at a high speed. Thus a number of charges were carried round and round the circumference of a circle, simulating a current in a circular loop of wire. A sensitive compass needle was suspended near the center of the revolving disk and its deflection used to determine the magnetic field produced. It was observed that the magnetic field was the same as if an equivalent number of coulombs per second were flowing in a wire around the circumference of the disk.

11. A somewhat similar, though less direct, experiment can be done with one of the Van de Graaff spheres. The current to the sphere is clearly carried by the mechanical movement of the silk belt, but it can be carried away from the sphere by a wire as an ordinary conduction current and its magnetic effect shown by running it through an ordinary galvanometer.

12. It is also possible to show by direct experiment that the current carried by the moving ions in electrolysis or by the moving ions in a gas discharge produces a magnetic field. The currents in both these processes are ordinarily rather small so that somewhat special arrangements are necessary to show their magnetic effect. To show the effect of an electrolytic current, glass tubing can be bent in a spiral and filled with an electrolyte. The current flowing between electrodes at the end of the spiral then is equivalent to a current flowing in a solenoid of a few turns, and its magnetic effect on a compass needle mounted in the spiral can be seen clearly. If the spiral is mounted with its axis pointing east and west, the needle will swing toward that direction when the current is turned on. A similar experiment with a solenoid might be done with a gas discharge, but it is possible to get enough current in a neon-lamp discharge in a straight tube to affect a compass needle. The current in such a tube is evidently equivalent to that flowing in a straight wire and affects a compass needle accordingly.

Force on a Current in a Magnetic Field. The Motor Law

13. In Chap. XIV, we described the repulsions and attractions between conductors carrying currents and between a cathode-ray beam and the current in a solenoid. We pointed out that the

introduction of the ideas of magnetism and magnetic fields would help us to handle such problems. So far, we have talked only about the magnetic fields produced by electric currents, not about the effects of such fields on currents. But there must be such effects. We saw that one current repelled another, and we know that if a current exerts a force on a magnet through the agency of a magnetic field, then, by Newton's third law, the magnet must exert a force of reaction on the current, also through the agency of the magnetic field. Many examples of this effect can be shown in lecture demonstrations. We shall come to the point at once and state the result which can then be applied to specific cases later. We shall call this law of force between an electric current and a magnetic field the *motor law.*

14. The force acting upon each element of a current-carrying conductor in a magnetic field is proportional to the strength of the current, the strength of the field, the length of the element, and the sine of the angle between the direction of the current and the direction of the magnetic field. Furthermore, the factor of proportion, if all the preceding quantities are expressed in m.k.s. units, proves experimentally to be the permeability μ which we have already encountered. We can, therefore, state the law as follows:

The force acting upon each element of length of a current-carrying conductor in a magnetic field is given by

$$\Delta F = I\mu H \Delta L \sin\theta \qquad (7)$$

where ΔF is the force, ΔL the element of length, I the current, H the magnetic field, μ the permeability, and θ the angle between the direction of the current and the direction of the magnetic field. The force is perpendicular to the plane determined by the current element and the magnetic field. It is out from this plane if θ measured from ΔL to H is counterclockwise; into this plane if θ is clockwise. If the current is measured in amperes, the magnetic field in ampere turns per meter, and ΔL in meters, the force is in newtons. An example is given in Fig. 149.

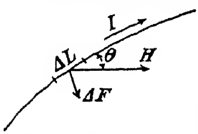


FIG. 149.—The force on a current in a magnetic field. *I* rotates clockwise into *H* and therefore the force ΔF is into the paper.

15. Besides the method used in the statement of the law, there are two other ways of determining the direction of the motor force. One is the so-called right-hand rule. According to this rule, if

the index finger is pointed in the direction of the current and the middle finger in the direction of the magnetic field, then the thumb jutting out at right angles to the plane determined by these two fingers points in the direction of the force. A second, less "rule-of-thumb," way of remembering this is to think of the lines of magnetic force. There is a certain distribution of lines of magnetic force before the current-carrying wire is introduced. The current in the wire itself produces a magnetic field which is represented by more lines of force. On one side of the wire, the field of the current is in the same direction as the original field so that the resultant field has more lines of force than before. On the other side of the wire, the field from the current is in the opposite direction to the original field so that the resultant field is less than before and is represented by fewer lines of force. In a plane perpendicular to the wire, the magnetic field is as shown in Fig. 150(*c*). It is often convenient to think of lines of force as having real physical existence, occupying space, and exerting pull along their length and pressure across it. In these terms, it is easy to see that the crowded lines of force above the wire will push it down. Whether the student cares to think of the lines of force as concretely as this or not does not matter. He will certainly find it convenient to remember that a current-carrying conductor will always try to move from a region of many lines of magnetic force toward one of few.

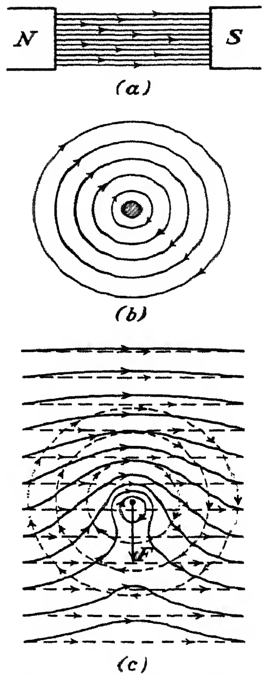


FIG. 150.—Modification of a magnetic field by a current-carrying wire. (*a*) The initial field. (*b*) The field around a wire with the current into the paper. (*c*) The resultant field. The wire tends to move from the stronger into the weaker field.

Magnetic Induction

16. The product μH which enters into the preceding equation for the motor law is so important that it is given a name of its

own. It is called the *magnetic induction* and is usually designated by B. Like H, it is a vector quantity. A current produces a field of magnetic induction, and we can draw lines of magnetic induction. We are going to encounter many relations in the forthcoming chapters that involve the induction, and the equations expressing them might equally well be written in terms of μH or of B. Usually, we shall write out the product μH explicitly as we have in the motor law above. In terms of B, the motor law is

$$\Delta F = BI\Delta L \sin \theta. \tag{8}$$

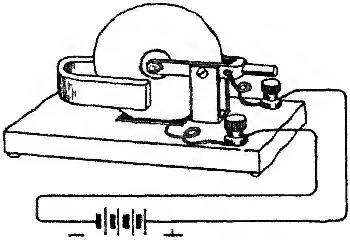


FIG. 151.—Barlow's wheel. The magnetic field is perpendicular to the plane of the disk and the current flows vertically down along a radius of the disk.

Barlow's Wheel

17. The sideways thrust of a magnetic field on a current-carrying conductor is the force that is used in all practical electric motors. In this paragraph, we shall describe an extremely simple type of motor working on this principle. A motor of this type would not suffice to pull the Pennsylvania Railroad trains between New York and Washington at 80 or 90 mi./hr. In fact, it is hardly more than a toy; but it is an application of the motor principle in so simple a form that it can very readily be understood. As is shown in Fig. 151, Barlow's wheel is merely a disk of copper mounted so that it can turn on a horizontal axis with its lower edge dipping into a trough of mercury. A horseshoe magnet is mounted so that the lower part of the disk is between the two poles. Thus there is a magnetic field in this region perpendicular to the surface of the disk. The positive terminal of a battery is now connected to the axle, and the negative terminal to the mercury trough so that the current flows through the axle out along a radius of the wheel and through the mercury back to the battery. Consequently, in the region between the poles of the magnet, there is a current flowing vertically in a horizontal magnetic field. There is, therefore, a force acting perpendicular to both the current and the field, and this force will tend to turn the disk. If the direction of the current is reversed, the wheel rotates in the other direction.

Current-measuring Instruments

18. The only method of measuring current that we have described so far depended on the electrodeposition of silver in electrolysis. This involved chemical weighing and at best only gave an average current over a period of time. Almost all the familiar instruments for measuring both currents and voltages depend on the electromagnetic effects of currents. The principal exception is a class of a-c instruments that depend on the heating effect of a current. The electromagnetic effects can be used in different ways as follows:

1. The current can be measured by sending it through a fixed coil and observing the effect of its magnetic field on a small magnet suspended so that it is free to move.

2. The current can be sent through a movable coil suspended in the magnetic field of a permanent magnet so that the "motor" force of the magnetic field on the current-carrying coil causes a motion which is a measure of the current.

3. The current can be sent through two coils, one of which can move. This can be thought of as differing from 2 only by having the current itself produce the magnetic field in which it moves. This type of instrument is particularly useful for a-c measurements.

4. Still a fourth type of instrument uses no permanent magnets but does use iron. It can be thought of as a combination of 2 and 3. The current to be measured flows through a coil or coils that magnetize pieces of iron. One of these is fixed and one movable. The magnetic force set up between these two pieces of iron is dependent on the current flowing. As in the other instruments, movement occurs until the restoring force of a spring or suspension is just equal to the magnetic repulsion. An instrument of this type can be used for either direct or alternating currents.

19. There are two standard ways of observing the movements of all four types of instruments and indeed of many other physical instruments. One, used on the more rugged and portable instruments, is to mount a pointer on the moving part and read the position of the pointer on a scale. The other is to mount a small mirror on the moving part and read the reflection of a fixed scale in it or shine a spot of light on it and observe the deflection of the reflected light beam on a fixed scale as the mirror turns.

Current-measuring Instruments. Examples

20. We shall describe in some detail examples of each of the first three types of instrument cited in Par. 18.

1. *The Tangent Galvanometer.* In this type of instrument, the current flows in a circular coil, sometimes of only one turn and of rather large diameter, perhaps 30 cm. At the center of this coil is a compass needle whose position can be read on a calibrated scale. The instrument has to be oriented in a vertical north-south (magnetic) plane so that when no current is passing through the coil the compass needle points north in the plane of the coil. When current passes through the coil, it produces a magnetic field at right angles to the plane of the coil. The needle then turns until the force of the earth's field tending to turn it back to the north is just equal to the force of the magnetic field of the current tending to turn it in an east-west direction. Or putting it another way, the compass needle points in the direction of the vector sum of the magnetic field of the earth and the magnetic field of the current. It is quite easy to show mathematically that the tangent of the angle through which the compass needle turns is proportional to the current flowing through the coil. It is this property of the instrument that gives it its name.

21. The advantage of this type of instrument is that deflection of the needle depends only on the strength of the earth's field, the dimensions of the coil, and the magnitude of the current flowing. The earth's field is quite easily measured, and therefore this instrument gives a direct measurement of current in absolute units. Other types of instrument give current measurements that are always relative to some other current. But the clumsiness of a tangent galvanometer makes it undesirable for ordinary work.

2. *D'Arsonval or Moving-coil Galvanometer*

22. In this type of instrument, a small rectangular coil of many turns of fine wire is mounted between the poles of a permanent magnet. To improve the magnetic efficiency, a cylinder of iron is mounted so that it fills the space inside the coil. The vertical sides of the coil move in annular slots between this core and the poles of the magnet. In the more sensitive instruments, the coil is suspended by a fine wire, usually a strip of phosphor bronze, so that the instrument has to be carefully leveled before it is used. In the less sensitive instruments, the coil is mounted in bearings

much like those which support the balance wheel of a watch. In either case, the coil hangs with its plane parallel to the lines of magnetic force when there is no current flowing. When a current flows, it goes downward in the wires on one side of the coil and upward in the wires on the other side of the coil so that the motor force of the magnetic field on the current is horizontal on each side but in opposite directions. The arrangement is shown in Fig. 152. The currents in the top and bottom of the coil are parallel to the magnetic field so that there is no force on them. The resultant effect is to turn the coil as shown in the figure. This

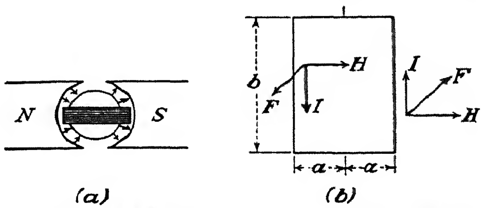


FIG. 152.—D'Arsonval galvanometer. (a) Schematic diagram of the coil of the galvanometer in a radial magnetic field. (b) The plane of the coil showing the directions of the magnetic field H, the current I, and the motor force F. They are mutually perpendicular.

turning force is opposed in one case by the rigidity of the wire suspension and in the other case by a light spring similar to the hair spring of a watch. These restoring forces become greater the greater the rotation (an example of Hooke's law, see Chap. II, Par. 31). The turning moment is proportional to the current so that the coil turns until the turning force is just balanced by the elastic restoring force of the suspension or spring. The angle through which the coil turns before reaching this position of equilibrium is a measure of the current.

23. More specifically, if the dimensions of the coil are as shown in Fig. 152(b) and it has n turns, the force on each side of the coil is $IBnb$ where B is the magnetic induction $= \mu H$. The turning moment of each of these forces about the axis is $IBnba$, and the total turning moment is twice this, or $IBn2ba$. It is interesting to notice that this expression can be written $IBnA$ where A is the area of the coil so that it makes no difference what the dimensions of the coil are as long as the area is the same.

24. The advantages of instruments of this type are several. The suspension type can be made extremely sensitive, responding

in some forms to currents as small as 10^{-9} amp. The pivot instruments are remarkably rugged even when sensitive enough to measure currents as small as 10^{-6} amp. In both types, the deflection is proportional to the current over the whole range, which is of great convenience. The only disadvantage is that the deflection depends not only on the current, but also on the field of the magnet and the elastic properties of the suspension or spring, and these are hard to evaluate in absolute terms. For practical purposes, this is immaterial since the instrument can always be calibrated in terms of the same current flowing through some other instrument. Probably 95 per cent of the galvanometers, ammeters, and voltmeters in use for the measurement of direct currents are of this type. They cannot be used for measuring alternating currents.

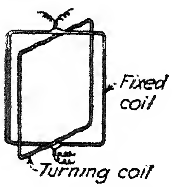


FIG. 153.—Electrodynamometer.

3. *Electrodynamometer*

25. In instruments of this type, a stationary coil is substituted for the permanent magnet. The current to be measured flows through both coils. In one arrangement, the coils are mounted at right angles to each other as shown in Fig. 153. There is a turning moment on the moving coil as in the permanent magnet instrument; but now if the current is reversed in the moving coil, it also reverses in the fixed coil, reversing the magnetic field. Consequently, the direction of the deflection remains the same. An instrument of this type consequently can be used for measuring alternating currents as well as direct.

4. *Iron core instruments.* Many a-c instruments are of this type, but the particular design of any given instrument involves a rather more complete understanding of the magnetization of iron by electric currents than we have given. We shall make no attempt to discuss an instrument of this type.

SUMMARY

Ampère's law for the contribution ΔH to the magnetic field at a point P by an element of current-carrying conductor ΔL is

$$\Delta H = \frac{I \Delta L \sin \theta}{4\pi r^2}$$

where I is the current, r is the distance from ΔL to P, and θ is the

angle between ΔL and r. H is perpendicular to the plane determined by r and ΔL, and its sense can be obtained by the right-hand rule.

Ampère's law applied to the magnetic field set up by an infinite straight wire shows that the lines of force are circles enclosing the wire and that the field at P distant r from the wire is $H_p = I/2\pi r$. Ampère's law applied to a circular loop confirms the experimental result given in the last chapter. The magnetic field at the center of a plane circular coil of radius a and of n turns is $H = nI/2a$, perpendicular to the plane of the coil. The work done in moving a unit positive pole once around a current-carrying conductor is I joules regardless of the path followed. This result is used to show that the magnetic field inside an infinitely long solenoid is $H = nI$ where n is the number of turns per unit length of the solenoid.

Rowland showed that static charges revolving on the rim of an insulating disk produced a magnetic field at the center of the disk equivalent to that produced by current in a circular loop of wire. The currents in electrolysis and in gas discharges also produce magnetic fields.

The force acting on each element of length ΔL of a current-carrying conductor in a magnetic field H is

$$\Delta F = I\mu H\Delta L \sin \theta$$

where μ is the permeability and θ is the angle between ΔL and H. The force is perpendicular to the plane determined by H and ΔL. This law is called the *motor law* and is the principle upon which all practical electric motors depend for their operation. The magnetic induction B is equal to μH.

Most instruments for the measurement of currents and voltages depend on Ampère's law, the motor law, or both. The most generally used instrument is the moving-coil galvanometer in which the current to be measured runs through a plane coil mounted on pivots between the pole pieces of a permanent magnet. The field of the magnet exerts a torque on the current in the coil which turns until this torque is just equal to the elastic restoring torque of the spring mounting.

ILLUSTRATIVE PROBLEMS

1. If the high-tension wires in a transmission line carry 1,000 amp. and are 2 m. apart, what is the force between them per meter of wire?

The magnetic field at a perpendicular distance r from a long straight wire carrying a current of I amp. is, Eq. (2), page 327,

$$H_p = \frac{I}{2\pi r}.$$

In this problem, the field at one wire due to the other is thus

$$H_p = \frac{1{,}000 \text{ amp.}}{2\pi \times 2 \text{ m.}}.$$

The force ΔF on a wire carrying a current of I amp. in a magnetic field of strength H amp. turns/m. is, Eq. (7), page 334,

$$\Delta F = I \Delta L \mu_0 H \sin \theta$$

where μ_0 is the permeability of empty space and θ is the angle between the current I and the magnetic field H. Since, in the transmission line, the field due to one wire current is circular as in Fig. 150(b), it will be at right angles to the second wire current, since this latter current would also be perpendicular to the paper in Fig. 150(b). Hence, $\theta = 90°$ and $\sin \theta = 1$. The second wire current is also 1,000 amp.

Since we are to find the force per meter of wire, $\Delta L = 1$ so that

$$\begin{aligned}\Delta F &= 1{,}000 \text{ amp.} \times 1 \text{ m.} \times 4\pi \times 10^{-7} \frac{\text{newton}}{\text{amp.}^2} \times \frac{1{,}000 \text{ amp.}}{2\pi \times 2 \text{ m.}} \times 1 \\ &= 0.1 \text{ newton.}\end{aligned}$$

2. A magnetic dipole of moment 4×10^{-7} newton-m.^{2}/amp. turns is placed at the center of a solenoid 1 m. long with 1,000 turns, carrying a current of 0.5 amp. What torque acts on the dipole when its axis is at right angles to that of the solenoid?

The magnetic field at the center of a solenoid is, Eq. 6, page 332,

$$H = nI = \frac{NI}{l}$$

where N = 1,000 turns = the total number of turns.
I = 0.5 amp. = the current in the solenoid.
l = 1 m. = the length of the solenoid.

Therefore,

$$H = \frac{1{,}000 \text{ turns} \times 0.5 \text{ amp.}}{1 \text{ m.}} = 500 \frac{\text{amp. turns}}{\text{m.}}.$$

This field is parallel to the axis of the solenoid of Figs. 138 and 147.

The torque on a magnetic bar in a magnetic field is, Eq. (6), page 320,

$$T = HM \sin \theta$$

where $H = \dfrac{500 \text{ amp. turns}}{\text{m.}}$ = the strength of the magnetic field.

$M = 4 \times 10^{-7} \dfrac{\text{newton-m.}^2}{\text{amp. turns}}$ = the moment of the magnet.

$\theta = 90°$ = the angle between the axis of the magnet and the direction of the magnetic field.

Therefore,

$$T = 500 \frac{\text{amp. turns}}{\text{m.}} \times 4 \times 10^{-7} \frac{\text{newton-m.}^2}{\text{amp. turns}} \times 1$$
$$= 2 \times 10^{-4} \text{ newton-m.}$$

3. What is the strength of the current in a tangent galvanometer of 2 turns of 32 cm. radius at a place where the horizontal intensity of the earth's magnetic field is 13 amp. turns/m., if the deflection of the needle is 30°?

The strength of the magnetic field at the center of the coil is, Eq. (4), page 329,

$$H = \frac{nI}{2a}$$

where H = ? = the strength of the magnetic field at the center of the coil.

n = 2 turns = the number of turns of the coil.

I = ? = the current in the coil.

a = 0.32 m. = the radius of the coil.

Therefore,

$$H = \frac{2 \text{ turns } I}{2 \times 0.32 \text{ m.}} = \frac{I}{0.32} \frac{\text{turns}}{\text{m.}}$$

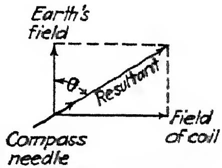


FIG. 154.—Magnetic fields acting on a compass needle in a tangent galvanometer. Tan θ = field of coil earth's field.

Now this magnetic field is at right angles to the plane of the coil. The plane of the coil is parallel to the earth's magnetic field. The compass needle will point in the direction of the resultant of these two perpendicular magnetic fields as in Fig. 154. The tangent of the angle θ between the direction of the earth's field and the compass needle is

$$\tan \theta = \frac{\text{field of coil}}{\text{earth's field}}.$$

In this problem $\theta = 30°$, the field of the coil is $\dfrac{I}{0.32} \dfrac{\text{turns}}{\text{m.}}$, and the earth's field is 13 amp. turns/m. Therefore,

$$\tan 30° = \frac{\dfrac{I}{0.32}\,\dfrac{\text{turns}}{\text{m.}}}{13\,\dfrac{\text{amp. turns}}{\text{m.}}}$$

$$\frac{1}{\sqrt{3}} = \frac{I}{13 \times 0.32}\,\frac{1}{\text{amp.}}$$

$$I = \frac{13 \times 0.32}{\sqrt{3}} \text{ amp.} = 2.40 \text{ amp., current in the coil.}$$

4. A square coil hangs in a region between two magnetic poles where the magnetic field strength is 10^4 amp. turns/m. If the length of one side of the coil is 1.5 cm. and if there are 100 turns in the coil, what is the torque on the coil when a current of 10^{-6} amp. exists in it?

The torque on a galvanometer coil is, *cf.* Par. 23,

$$T = I\mu_0 HnA$$

where $I = 10^{-6}$ amp. = the current in the coil.

$$\mu_0 = 4\pi \times 10^{-7}\,\frac{\text{newton}}{\text{amp.}^2} = \text{the permeability of empty space.}$$

$$H = 10^4\,\frac{\text{amp. turns}}{\text{m.}} = \text{the magnetic field strength.}$$

n = 100 turns = the number of turns of the coil.

$A = (1.5 \text{ cm.} \times 10^{-2} \text{ m./cm.})^2$ = the area of the coil.

Therefore,

$$T = 10^{-6} \text{ amp. } 4\pi \times 10^{-7}\frac{\text{newton}}{\text{amp.}^2} \times 10^4\,\frac{\text{amp. turns}}{\text{m.}} \times 100 \text{ turns} \times (1.5 \times 10^{-2} \text{ m.})^2$$

$$= 2.83 \times 10^{-10} \text{ newton-m.}$$

Here the turns^2 in the numerator are either considered a pure number or turn^2 is associated with amp.^{2} in the denominator and they are canceled out.

PROBLEMS

1. At what distance from a long straight wire carrying a current of 25 amp. is the magnetic field strength equal to 8 amp. turns/m.?

2. Two straight parallel wires 1 cm. apart each carry 3 amp. of current. What is the force of attraction between a 30-cm. section of one wire and the other wire?

3. How much current is flowing in the third rail of an electric railway that runs in a north and south direction and causes a deflection of 45° in a compass needle held 30 cm. above the center of the rail; taking the strength of the horizontal component of the earth's magnetic field as 16 amp. turns/m.?

4. A straight horizontal wire runs parallel to the magnetic meridian at a point where the horizontal component of the earth's magnetic field is

13 amp. turns/m. A compass needle 10 cm. above the wire stands at an angle of 45° with the magnetic meridian. How much current is there in the wire?

5. A magnetic dipole of moment 3×10^{-7} weber-m. is placed horizontally 20 cm. above a straight horizontal wire with its axis at an angle of 45° to the wire. What torque acts on the dipole when there is a current of 100 amp. in the wire?

6. In a manner analogous to that of Par. 5, find the magnetic field on the axis of a circular coil of n turns at a distance f from the center of the coil. Remember that the magnetic field is a vector so that the contribution for each length ΔL should be resolved into components parallel and perpendicular to the axis of the coil. The components contributed by the different sections of wire ΔL are then added together.

7. A pair of Helmholtz coils consists of two coils in which the current circulates in the same sense mounted with their planes parallel a distance apart equal to the radius of one of the coils. They give a fairly uniform magnetic field near the point on the axis midway between the coils. If the coils have 20 turns of 30 cm. radius carrying 10 amp., what is the magnetic field at the above point?

8. How many turns per meter must a solenoid have in order to have the same magnetic field at the center as a coil of 20 turns of 2 cm. radius if the current in the coil is three times that in the solenoid?

9. One advantage of a solenoid is that it produces a nearly uniform magnetic field inside. A solenoid has 1,000 turns/m. carrying a current of 10 amp. What force will act on 1 cm. of a beam of 10^{12} electrons/cm. moving with a velocity of 10^6 m./sec. along a line making an angle of 30° with the axis of the solenoid?

10. What must be the linear charge density on Rowland's disk turning sixty-one times a second if the magnetic field at its center is to be 0.0192 amp. turn/m.?

11. The rim of Rowland's disk has a linear charge density of 10 microcoulombs/m. and rotates 61 r.p.s. The plane of the disk is vertical and in the magnetic meridian where the horizontal component of the earth's magnetic field is 13 amp. turns/m. What will be the deflection of a small compass needle mounted at the center of the disk?

12. In Barlow's wheel, a uniform magnetic field extends from a radius of 2 to 6 cm. with a strength of 2×10^5 amp. turns/m. If a current of 40 amp. flows in the disk, what is the force acting to rotate the disk? What is the resultant moment of this force?

13. A circular coil of seven turns of wire 18 cm. in diameter carries a current of 3 amp. What will be the deflection of a magnetic needle at the center of this coil if it is placed with its plane vertical and in the magnetic meridian at a point where the earth's horizontal component of magnetic field strength is 13 amp. turns/m.?

14. A D'Arsonval galvanometer has a coil of area 2 sq. cm. with 100 turns. The magnetic induction is 2 newtons/m. amp. turns. What torque will be produced in this coil by a current of 1 microamp.?

CHAPTER XVII

THE LAWS OF FLOW OF ELECTRIC CURRENTS

1. In Chap. XIII, we introduced the idea of electric currents as streams of charged particles moving under the influence of potential differences. Subsequently, we took the existence of currents for granted and investigated the magnetic fields they produce and the effect of magnetic fields on them. In the present chapter, we return to currents and potential differences for a quantitative study. We shall consider the relations between potential differences and currents in solid conductors and some other closely related phenomena. Before embarking on these topics, we shall review the question of units.

2. We first introduced the coulomb as the unit of charge by giving an arbitrary value to the factor of proportion in Coulomb's inverse-square law. Then we introduced the volt as a unit of potential difference suggested naturally by the universal early use of the Daniell cell and defined precisely in terms of the Weston standard cell. We then defined the coulomb in terms of the volt and the joule, the unit of energy already established in mechanics, and thereby justified the manner in which it had been previously introduced. The unit of current, called the ampere, then follows naturally as the flow of one coulomb per second.

3. Granted the joule as the unit of energy, it is evident that a unit of only one of the three electrical quantities needs to be defined independently. The ampere, the coulomb, or the volt could be defined and the other two definitions derived from it and the joule. We have chosen the volt because it is a more or less natural unit. For purposes of exact legal standards, it is better to use the ampere and to define it in terms of the force between two parallel wires.

Electrical Resistance

4. In the early chapters of this book, we discussed the kinetic theory of matter and saw that all matter was composed of small

particles in continual chaotic motion. This chaotic motion will interfere with the steady motion of a particle of small dimensions through matter. The passage of ions through a solution or through a gas or the passage of electrons through a gas or a solid is continually retarded by collisions with the uncharged molecules that are in thermal motion. The ion is continually accelerated by the electric field and increases its velocity in the direction of the field. But at a collision of the ion with a molecule, the laws of momentum and energy hold just as they do in a collision between two billiard balls. The ion therefore gives up some of its kinetic energy to the molecule, and its momentum changes in both magnitude and direction. The result is that the motion of the ion toward the electrode is retarded and the kinetic energy of the uncharged molecule is increased. In other words, the conducting material, be it gaseous, liquid, or solid, offers resistance to the passage of current, and the work done in overcoming this resistance goes into heating the conductor. The amount of resistance offered and, consequently, the electric field strength necessary to produce a given current in a given volume of a conductor depend on the nature of the material and on its temperature. Thus the potential drop that must be set up to drive a current between two points of a conductor depends on the length of the conductor between these points, the cross section and material of the conductor, the temperature, and the current.

Ohm's Law. The Unit of Resistance

5. Fortunately, the law of dependence of current on potential difference is a very simple one for the case which is of most use, the flow of currents through metallic conductors at ordinary temperatures. In this case, the current flowing is directly proportional to the potential difference. This is known as Ohm's Law and may be formally stated as follows:

According to Ohm's law, for any given conductor, the current flowing is proportional to the difference of potential between the ends of the conductor.

6. The ratio of the fall of potential across any conductor that forms part of an electric circuit to the current flowing through it is called the resistance of that conductor. If this ratio is designated by R and the current and potential drop by I and E, respectively, Ohm's law can be stated mathematically as

$$I = \frac{E}{R} \qquad \text{or} \qquad E = RI. \tag{1}$$

7. Obviously this equation can be used to define a unit of resistance. The unit of resistance that we shall use is the unit in the practical system. It is called the ohm and is defined by the following statement:

A conductor has a resistance of one ohm if there is a difference of potential of one volt between its ends when a current of one ampere is flowing through it.

A Complete Circuit

8. In any ordinary d-c circuit, there is a steady flow of charge around a closed path under the influence of the e.m.f. of either a battery or a dynamo. Such a circuit with a battery as the

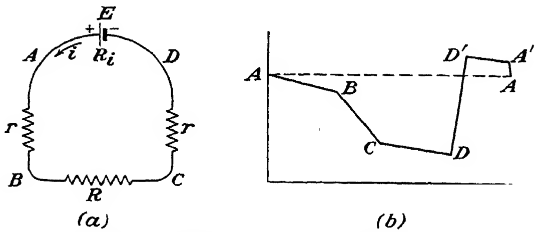


FIG. 155.—(*a*) Schematic diagram of a simple circuit. E is the e.m.f. of the battery whose internal resistance is R_i, r and r are the resistances of the connecting wires, and R is the resistance of the device operated by the battery. (*b*) Potential drops around the circuit of Fig. 155(*a*).

source of potential is shown in Fig. 155(*a*); it may include an incandescent lamp, a motor, or any other device operated by electric current. Any such device will offer resistance to the passage of current and can be represented by the circuit diagram of Fig. 155(*a*) where R is the resistance of the device. The connecting wires and the electrolyte of the battery itself also have resistance. It is instructive to follow the variation of potential around the circuit. Suppose the battery is a simple copper-zinc-sulfuric acid cell such as we have described. If we start at the positive terminal A, the potential falls off slightly in the connecting wire to B, one terminal of the resistance R. Between B and C, because of the resistance, there is a larger fall of potential. Between C and D, the negative

terminal of the cell, there is again a slight fall of potential through the connecting wire. Between the zinc electrode D and the adjacent solution D', there is a sharp rise in potential caused by the chemical reaction, and it is this rise in potential which gives the force that pushes the charges around the circuit, making the current flow. Some of the energy supplied by the chemical reaction at this point is used in moving the ions through the solution, *i.e.*, in overcoming the internal resistance of the cell, so that there is a fall in potential between D' and A', the point in the solution just adjacent to the copper electrode. Finally, there is a sudden drop in potential in going from the solution to the copper electrode A. This completes the circuit, bringing us back to our starting point. As we saw in our discussion of this cell, the changes in potential at the electrode surfaces are characteristic of the electrodes and the solution, and the difference between these two discontinuities in potential is the e.m.f. of the cell. In our diagram, it is the difference between DD' and AA'. If we call this E, then it is clear that E must be equal to the sum of all the other potential drops around the circuit. Calling the resistances of the connecting wires r and r', the internal resistance of the cell R_i, and the current I, we have from Ohm's law

$$E = I(R + R_i + r + r').$$

This result obtained by a detailed consideration of this particular closed circuit can be proved in more general form for any closed circuit containing any number of resistances and e.m.fs. For such a circuit, it becomes:

In any closed circuit, the sum of the electromotive forces must equal the sum of the products of the respective currents and resistances in the various parts of the circuit. In mathematical language,

$$\Sigma E = \Sigma IR. \tag{2}$$

Evidently Ohm's law is a special case of this more general statement.

9. The law just stated is usually known as Kirchhoff's Second Law. The first of Kirchhoff's laws is merely an expression of the fact that charge will not accumulate at any point in a closed circuit. If this is so, then the amount of current flowing in to any point must be equal to that flowing out. This is Kirchhoff's first law, which may be formally stated as follows:

The currents flowing to any point of a conductor being taken as positive and those flowing away from the point being taken as negative, the sum of the currents at any point is equal to zero. Mathematically,

$$\Sigma I = 0. \tag{3}$$

Resistances in Series and in Parallel

10. If several resistances R_1, R_2, . . . are connected in series as shown in Fig. 156(a), it is clear that the current flowing through

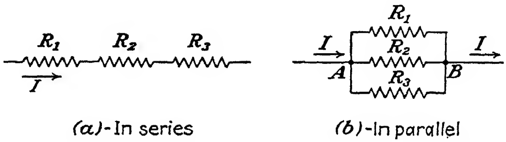


FIG. 156.—Resistances connected in series and in parallel.

each of them is the same, as otherwise there would be an accumulation of charge between them. Therefore, the potential drops across them must be IR_1, IR_2, ... The total potential over the whole series must be the sum of these, or

$$I(R_1 + R_2 + R_3 + \cdots).$$

This is the same potential drop that would be caused by a single resistance of value $R = R_1 + R_2 + R_3 + \cdots$. Consequently, we see that the resistance of a set of resistances connected in series is equal to the sum of the separate resistances.

11. If, on the other hand, a number of resistances are connected in parallel as shown in Fig. 156(b), the current will split, some of it going through each resistance. But the potential difference V between the points A and B is a definite quantity so that the IR drop must be the same over all branches of the circuit between these points. If we let I_1, I_2, . . . be the currents flowing through the various branches of the circuit, then we must have

$$V = I_1R_1 = I_2R_2 = I_3R_3 \tag{4}$$

or

$$\frac{V}{R_1} = I_1, \qquad \frac{V}{R_2} = I_2, \qquad \frac{V}{R_3} = I_3 \tag{5}$$

Adding these equations, we have

$$V\left(\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots\right) = I_1 + I_2 + I_3 \cdots \tag{6}$$

but $I_1 + I_2 + I_3 + \cdots = I$, the total current. Therefore, if R is the single resistance equivalent to this set of resistances connected in parallel,

$$V = RI \qquad \text{or} \qquad V\left(\frac{1}{R}\right) = I.$$

Therefore, comparing this with Eq. (6), we see that

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \tag{7}$$

which is the law giving the resistance equivalent to a number of resistances connected in parallel.

Ammeters and Voltmeters

12. At the end of the last chapter, we discussed the principles on which various types of current-carrying instruments worked. We now want to explain the particular way in which a galvanometer is used to measure a voltage or a current. In making any physical measurement, it is desirable to have the process of measurement affect the conditions that are being measured as little as possible. Thus, if we are measuring the potential drop between two points of a circuit, we want to do it without altering the potential drop between these points; or if we are measuring the current through a part of a circuit, we do not want the presence of the measuring instrument to increase or decrease the current. It is for this reason that it is necessary to use a galvanometer in quite different ways for the two types of measurement, although the deflection of the galvanometer coil itself always depends on the current through it and, therefore, on the potential drop between its terminals.

13. To measure a current, we want to find out how many units of charge per second are passing through any plane intersecting the conductor. We want to cut a conductor at a point and insert there an instrument which will, so to speak, count the charges going by. If this instrument has a high resistance, it will cut

down the flow of current. What we want then is a low-resistance galvanometer connected in series with the circuit. This is essentially what any ammeter is. For convenience, the scale of deflections of the galvanometer is calibrated to read directly in amperes. Sometimes it is impractical to have the resistance of the coil of the galvanometer low enough to carry the whole of the current without appreciable heating. In such a case, only part of the current is sent through the coil, and the rest goes through a low resistance connected in parallel. Such a resistance is said to be shunted across the coil and is itself called a shunt. Such a shunt may be mounted inside the case of the instrument, as is usually done for ammeters measuring currents up to 15 amp., or may be outside the instrument, as is usually done for the measurement of very large currents.

14. A potential difference by definition involves two points. A voltage is to be measured between two different points in a circuit. Consequently, a voltmeter cannot be connected in series since this means merely inserting it at one point in the circuit. A voltmeter must be connected in parallel, shunted across the part of the circuit over which the potential drop is to be measured. Furthermore, the voltmeter must take so little current that the conditions in the circuit are not appreciably affected. The voltmeter must have a high resistance. As in the ammeter, this resistance does not necessarily have to be in the moving coil. It may be separate, in series with the coil, and mounted either inside the case of the instrument or separately. A galvanometer with a high resistance to be used as a voltmeter usually has its scale calibrated directly in volts.

15. It is clear from the foregoing paragraphs that the same galvanometer can be used to measure either current or voltage by using it with various auxiliary resistances. This is often done in practice.

The Potentiometer

16. Even a high-resistance voltmeter takes some current from the source of potential it is measuring. It is sometimes desirable for very accurate measurements or measurements of potential differences that are very feeble sources of power to have an instrument that supplies its own power. Such an instrument can be constructed very simply by the application of Ohm's law

in the manner schematically shown in Fig. 157. It is called a potentiometer. The battery sends a current through the resistance AB and thereby produces a potential drop through that resistance. In the simplest form of potentiometer, AB is a uniform straight wire of some material that has a fairly high resistance. The potential drop between A and any point P along the wire is then proportional to the length AP; it is, in fact, equal to $V \times AP/AB$ where V is the potential drop over the whole length of the wire AB as measured by a voltmeter. Now connect A to one of the two points between which the potential difference is to be measured, putting a galvanometer in series. Connect the other point to the sliding contact P. Adjust this contact until no current flows through the galvanometer. This can occur only when the potential difference being measured is just equal to the potential difference AP. The potential difference being measured is therefore $AP/AB \times V$. Since no current is flowing through the galvanometer, no power is being taken from the unknown potential difference and it is therefore unaffected by the measurement. The power for the measurement, for operating the voltmeter and maintaining the potential drop through AB, comes from the battery. Though all potentiometers are based on this principle, the simple straight resistance wire is usually replaced by coils and the contact is shifted by turning knobs on the outside of a box containing the coils. If the wiring diagrams of such instruments are studied, they are found to be equivalent to that of Fig. 157.

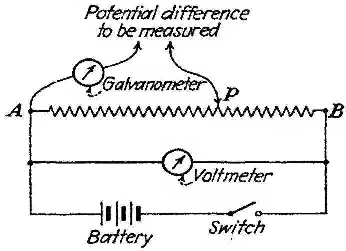


FIG. 157.—Schematic diagram of a potentiometer. The difference of potential to be measured is balanced against the drop of potential AP in the wire AB.

The Wheatstone Bridge

17. Resistance can be measured with sufficient accuracy for many purposes by using a voltmeter and ammeter. More accurate determinations can be made by the simple arrangement known as a Wheatstone bridge. This is probably familiar from

previous study as is much of the material of this chapter. It is included here for the sake of completeness and because it is a simple illustration of a branched circuit. The arrangement of resistances, battery, and galvanometer is shown in Fig. 158 where R_1, R_2, R_3 are known resistances and R_x is the resistance to be measured. One of the known resistances, say R_3, is adjustable. To measure the unknown resistance, R_3 is adjusted until the galvanometer shows that no current is flowing between B and C. This means that B and C must be at the same potential, and therefore the potential drop from A to B is the same as from A to C; also the potential drop from B to D must be the same as from C to D. But the current flowing in through A splits into two parts, one going through AB and the other through AC. Call these two currents I_1 and I_2, respectively, and the currents through BD and CD, I_x and I_3. Then if there is no current through the galvanometer, $I_1 = I_x$ and $I_2 = I_3$. Also the potential drops over the four branches of the circuit are I_1R_1, I_2R_2, I_3R_3, and I_xR_x from Ohm's law. Therefore,

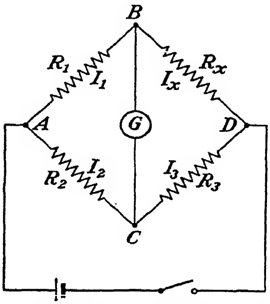


FIG. 158.—Circuit diagram for a Wheatstone bridge. If no current flows through the galvanometer, the points B and C are at the same potential.

$$R_1I_1 = R_2I_2 \tag{8}$$

$$R_xI_x = R_3I_3 \tag{9}$$

but

$$I_x = I_1 \qquad I_3 = I_2 \tag{10}$$

Therefore

$$R_xI_1 = R_3I_2. \tag{11}$$

Dividing this by the first equation, we get

$$\frac{R_x}{R_1} = \frac{R_3}{R_2} \tag{12}$$

or

$$R_x = \frac{R_1R_3}{R_2} \tag{13}$$

This last equation gives the unknown resistance in terms of the three known resistances.

Resistivity

18. In the opening sections of this chapter, we pointed out that the resistance of a conductor depended on its size and shape and the material from which it was made. Now that we have seen how a resistance is measured, we can investigate this question more closely. We shall confine ourselves to wires of uniform cross section. We can think of the wire as made up of a whole lot of sections of unit length connected in series. We can also think of each unit length as made up of a number of conductors each of unit cross section. Therefore, if the resistance of a unit length of the wire of unit cross section is ρ, then the resistance of unit length r will be given by

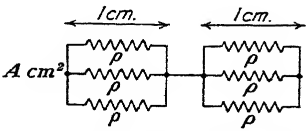


FIG. 159.—Schematic diagram of resistivity. The resistances in parallel must be added reciprocally and the groups in series be added directly to get the total resistance.

$$\frac{1}{r} = \frac{1}{\rho} + \frac{1}{\rho} + \frac{1}{\rho} + \cdots = \frac{A}{\rho}$$

or

$$r = \frac{\rho}{A}$$

if A is the cross section of the wire. If the length of the wire is l, then the resistance of the whole wire R is

$$R = lr = \frac{\rho l}{A} \tag{14}$$

The factor ρ in this equation is called the resistivity of the material of which the wire is made.

The resistivity of the material in a conductor is the ratio of the potential drop per unit length in the direction of the current to the magnitude of the current per unit cross section of the conductor.

The resistivity of a material is numerically equal to the resistance between two opposite faces of a cube of the material one meter on each edge.

TABLE 14.—RESISTIVITY

$$\rho = \frac{AR}{l}$$

Substance	ohm-m. at 20°C.
Aluminum	2.62×10^{-8}
Copper	1.77×10^{-8}
Gold	2.44×10^{-8}
Iron	10×10^{-8}
Lead	22×10^{-8}
Manganin	44×10^{-8}
Nichrome	100×10^{-8}
Platinum	10×10^{-8}
Silver	1.62×10^{-8}

Temperature Coefficient of Resistance

19. We have already pointed out that the passage of electrons through a metal is hindered by the chaotic thermal motion of the atoms of the metal. Consequently, it is not surprising to find that the resistance of most conducting materials increases with increasing temperatures. In materials such as carbon, where this is not true and the resistance decreases as the temperature rises, it is because the increasing motion of the atoms alters the structure of the material sufficiently to overcome the increased probability of collisions. As a matter of fact, the complete theory of the conduction of electricity through solids is very much more complicated than we have indicated and is by no means completely worked out as yet.

TABLE 15.—TEMPERATURE COEFFICIENTS OF RESISTANCE
Change of resistance per unit resistance per degree change in temperature
$R = R_0(1 + \theta T)$.

Substance	θ	Temperature range, °C.
Aluminum	3.8×10^{-3}	18–100
Copper	4.3×10^{-3}	18
Gold	4×10^{-3}	0–100
Lead	4.3×10^{-3}	18
Manganin	$2 \times 10^{-6} - 5 \times 10^{-5}$	20
Nichrome	1.7×10^{-4}	20
Platinum	3.8×10^{-3}	0–100
Silver	4.0×10^{-3}	0–100

The Heating Effect of a Current

20. It was also mentioned that some of the energy of motion of the electrons carrying a current was communicated to the atoms of the conductor with a consequent rise in its temperature. It is this effect, of course, that is used in the ordinary incandescent light bulb. It is found that the amount of energy which an electric current gives up as heat is proportional to the square of the current and to the first power of the resistance of the conductor carrying the current. As a matter of fact, we can show that this must be correct just from the definition of potential difference and resistance. Suppose that a potential difference of E volts is necessary to push a current of I amp. through a resistance of R ohms. By definition, this means that I coulombs of charge are being carried against the resistance every second and that E joules of work are being done for every coulomb. Therefore, the power being used is EI joules/sec. or EI watts. But $E = IR$ so that the energy expended per second is I^2R joules if I is in amperes and R in ohms. We may formulate this law as follows:

The rate at which energy is expended in any part of an electric circuit is equal to EI or I^2R watts, and the total energy expended is equal to EIt or I^2Rt joules, where E is the potential drop, I the current, R the resistance for the part of the circuit in question, and t the time.

21. Since the heating of a wire in a transmission line is a sheer waste of power, it is customary to transmit power at as low currents as possible. Large quantities of power are transmitted by using very high voltages but keeping the current down.

SUMMARY

Experiment shows that in metallic conductors the current flowing is proportional to the potential difference between the ends of the conductor. This is Ohm's law. The ratio of potential difference to current is called the resistance of a conductor. The unit of resistance is the ohm, which is a resistance through which a potential difference of one volt will drive a current of one ampere. Kirchhoff's laws for any electrical circuit include Ohm's law as a special case. They are as follows:

1. The amount of current flowing in to any point of a circuit is just equal to the amount flowing out.

2. In any closed circuit, the sum of the electromotive forces is equal to the sum of the potential drops around the circuit.

The resistance of a group of resistances connected in series is the sum of the separate resistances. The reciprocal of the resultant resistance of a group of resistances connected in parallel is the sum of the reciprocals of the separate resistances.

An ammeter is a low-resistance galvanometer inserted in series at some point in a circuit. A voltmeter is a high-resistance galvanometer connected in parallel between two points of a circuit. A potentiometer is a device for measuring the potential difference between two points of a circuit without taking any power from that circuit or affecting it in any way.

A Wheatstone bridge is an arrangement of resistances in a branched circuit which is used for the accurate measurement of unknown resistances. The resistivity of a material is numerically equal to the resistance between two opposite faces of a unit cube of the material. The resistance of a wire of length l and cross section A is $\rho l/A$ if ρ is the resistivity of the material of which the wire is made. In general, the resistance of a conductor increases with temperature.

Current flowing through a wire produces heat at the rate of I^2R watts.

ILLUSTRATIVE PROBLEMS

1. Two coils have a resistance of 50 ohms when connected in series and 12 ohms when connected in parallel. Find the resistance of each coil.

Paragraph 10 shows that two resistances connected in series have a total resistance equal to the sum of the separate resistances or

$$R = 50 \text{ ohms} = R_1 + R_2.$$

Equation (7), page 351, gives for the resistance R in parallel

$$\frac{1}{R} = \frac{1}{12 \text{ ohms}} = \frac{1}{R_1} + \frac{1}{R_2}.$$

Substitute into this equation the value $R_1 = 50$ ohms $- R_2$ from the series equation.

$$\frac{1}{12 \text{ ohms}} = \frac{1}{50 \text{ ohms} - R_2} + \frac{1}{R_2}.$$

Clearing of fractions gives

$$(50 \text{ ohms} - R_2)R_2 = 12 \text{ ohms} \times R_2 + 12 \text{ ohms } (50 \text{ ohms} - R_2).$$

Multiplying out the parentheses and transposing give

$$R_2^2 - 50 \text{ ohms} \times R_2 + 12 \times 50 \text{ ohms}^2 = 0$$

$$R_2 = \frac{50 \text{ ohms} \pm \sqrt{(50 \text{ ohms})^2 - 4 \times 12 \times 50 \text{ ohms}^2}}{2} = \frac{50 \pm 10}{2} \text{ ohms}$$

$$R_2 = 30 \text{ ohms or } 20 \text{ ohms}$$

$$R_1 = 50 \text{ ohms} - R_2 = 50 \text{ ohms} - 30 \text{ ohms} = 20 \text{ ohms}$$

or

$$R_1 = 50 \text{ ohms} - R_2 = 50 \text{ ohms} - 20 \text{ ohms} = 30 \text{ ohms}$$

so that the two resistances are 20 ohms and 30 ohms.

2. Two lamps of 60 and 180 ohms resistance, respectively, are connected in series to a constant difference of potential of 120 volts. Find the potential difference across the 60-ohm lamp.

The circuit is shown in Fig. 160. As in Prob. 1, the total resistance is the sum of the separate resistances or 240 ohms. From Eq. (1), page 348,

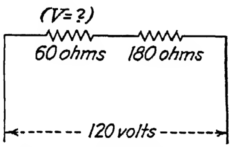


FIG. 160.—Two resistances of 60 and 80 ohms in series with a difference of potential of 120 volts.

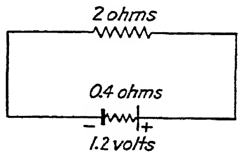


FIG. 161.—A simple cell in series with a 2 ohm resistance. Illust. Prob. 3.

$I = E/R = 120$ volts/240 ohms $= 0.5$ amp. Again using the same equation, we have $E = IR = 0.5$ amp. $\times$ 60 ohms $=$ 30 volts difference of potential across the 60-ohm lamp. The resistance of the connecting wires is neglected in these problems.

3. A single cell of e.m.f. 1.2 volts and internal resistance 0.4 ohm is connected to an external resistance of 2 ohms. What is (*a*) the current in the circuit, (*b*) the drop in potential due to resistance in the cell, (*c*) the difference of potential between the terminals of the cell, and (*d*) the difference of potential between the terminals of the external resistance?

In the circuit, Fig. 161, the total resistance is again the sum of the separate resistances connected in series. Therefore,

$$R = 2 \text{ ohms} + 0.4 \text{ ohm} = 2.4 \text{ ohms}$$

a. The current in the circuit is given by Eq. (1), page 348,

$$I = \frac{E}{R} = \frac{1.2 \text{ volts}}{2.4 \text{ ohms}} = 0.5 \text{ amp.}$$

The total resistance must be used.

b. The drop in potential due to any resistance is $E = IR$ so that the drop in potential due to resistance in the cell is $E = 0.5$ amp. $\times$ 0.4 ohm $=$ 0.2 volt.

c. The difference in potential between the terminals of the cell is made up of two parts, the e.m.f. of the cell of 1.2 volts produced by chemical action and the drop in potential due to resistance found in (*b*). If we go through the cell in the direction of the current in the circuit, the potential rises 1.2 volts owing to chemical action and falls 0.2 volt owing to resistance. The net difference of potential between the terminals is, therefore, 1.2 volts − 0.2 volt = 1.0 volt. The potential due to resistance must fall in the direction of the current, as it is this potential difference that drives the current through the resistance.

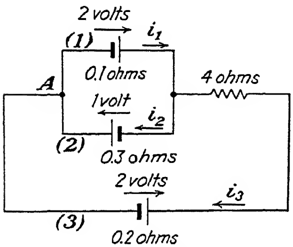


FIG. 162.—Three cells connected in a network. The arrows above each cell represent the normal direction of the current through each cell if the cell were connected to a simple resistance. The e.m.f. of each cell is given above the cell and its resistance below it.

d. The terminals of the cell and those of the external resistance are the same two points so that the difference of potential between the terminals of the external resistance is the same as that between the terminals of the cell, or 1.0 volt. The e.m.f. of the cell is completely used up in driving the current through the internal and external resistances, or 1.2 volts = 1.0 volt (external) + 0.2 volt (internal).

4. Three cells are connected as shown in Fig. 162. Find the current in each branch of the circuit.

Let the currents in the three branches be as shown by the arrows. Applying Eq. (3), page 350, to the point *A* gives

$$-i_1 + i_2 + i_3 = 0. \qquad (15)$$

Next we apply Eq. (2), page 349 to the three circuits formed when we take the cells in pairs. The circuit through (1) and (2) gives

$$2 \text{ volts} - 0.1 \text{ ohm} \times i_1 + 1 \text{ volt} - 0.3 \text{ ohm} \times i_2 = 0. \qquad (16)$$

The circuit through (1) and (3) gives

$$2 \text{ volts} - 0.1 \text{ ohm} \times i_1 - 4 \text{ ohms} \times i_3 - 2 \text{ volts} - 0.2 \text{ ohm} \times i_3 = 0. \qquad (17)$$

The circuit through (2) and (3) gives

$$-1 \text{ volt} + 0.3 \text{ ohm} \times i_2 - 4 \text{ ohms} \times i_3 - 2 \text{ volts} - 0.2 \text{ ohm} \times i_3 = 0. \qquad (18)$$

In each case, we have gone around the circuit in a clockwise direction. It must be remembered that potentials due to resistance are negative in the direction of the current. Equations (16), (17), and (18) are not independent as (17) − (18) = (16) which serves as a check of the equations. We now solve Eqs. (15), (16), and (17) simultaneously for i_1, i_2, and i_3. From Eq. (15), $i_3 = i_1 - i_2$. Substitute this in Eq. (17).

$$-0.1 \text{ ohm} \times i_1 - 4.2 \text{ ohms } (i_1 - i_2) = 0$$
$$-4.3 \text{ ohms} \times i_1 + 4.2 \text{ ohms} \times i_2 = 0$$

$$i_2 = \frac{4.3}{4.2} \times i_1. \qquad (19)$$

This value of i_2 is now substituted in Eq. (16).

$$3 \text{ volts} - 0.1 \text{ ohm} \times i_1 - 0.3 \text{ ohm} \frac{4.3}{4.2} \times i_1 = 0$$

$$3 \text{ volts} = \frac{1.71 \text{ ohms}}{4.2} \times i_1$$

$$i_1 = \frac{3 \text{ volts} \times 4.2}{1.71 \text{ ohms}} = \frac{42}{5.7} \text{ amp.}$$

This value of i_1 is now substituted in Eq. (19) to get

$$i_2 = \frac{4.3}{4.2} \times i_1 = \frac{4.3}{4.2} \times \frac{42}{5.7} \text{ amp.} = \frac{43}{5.7} \text{ amp.}$$

i_3 is now obtained from Eq. (15)

$$i_3 = i_1 - i_2 = \frac{42}{5.7} \text{ amp.} - \frac{43}{5.7} \text{ amp.} = -\frac{1}{5.7} \text{ amp.}$$

The minus sign indicates that the direction assumed by the arrow in Fig. 162 is reversed. Each current flows through each cell in the direction of its e.m.f.

5. A galvanometer of 100-ohms resistance with 150 divisions of 1 mm. each deflects 1 mm./milliamp. How much resistance must be put in series with this instrument to measure 150 volts with a full-scale deflection? What shunt is required to measure 15 amp. with full-scale deflection?

In order to get full-scale deflection, the current in the instrument must be 150 milliamp. The difference of potential when the instrument is used as a voltmeter is 150 volts. From Ohm's law,

$$R = \frac{E}{I} = \frac{150 \text{ volts}}{0.15 \text{ amp.}} = 1{,}000 \text{ ohms.}$$

1,000 ohms $= R + 100$ ohms since the added resistance is in series with the 100 ohms, of the galvanometer coil. $R = 900$ ohms.

The difference of potential between the terminals of the galvanometer with full-scale deflection is $E = IR = 0.15 \times 100 = 15$ volts. This is also the difference in potential across the shunt since the shunt is a resistance in parallel with the galvanometer. The current in the line divides, 0.15 amp. going through the galvanometer and 15.0 amp. $-$ 0.15 amp. $=$ 14.85 amp. going through the shunt. The resistance of the shunt is, therefore, $R = E/I = 15$ volts/14.85 amp. $=$ 1.01 ohms.

6. A potentiometer consists of a manganin slide wire of 0.2 mm.^{2} cross section mounted on a meter bar. A 6-volt storage battery of 0.09-ohm internal resistance supplies the power. How much resistance must be put in series with the slide wire in order to have 80 cm. balance 0.01 volt?

The resistance of 1 m. of manganin wire Eq. (14), page 355, is

$$R = \frac{\rho l}{A} = \frac{44 \times 10^{-8} \text{ ohm m.} \times 1 \text{ m.}}{0.2 \times 10^{-6} \text{ m.}^2}$$
$$= 2.20 \text{ ohms}$$

where 44×10^{-8} is the resistivity from Table 14. The resistance of 80 cm. of the wire is 0.8×2.20 ohms = 1.76 ohms, since the resistance is proportional to the length of the wire. The difference of potential across this 80 cm. is given as 0.01 volt so that the current in the wire and, hence, in the rest of the series circuit of Fig. 163 is

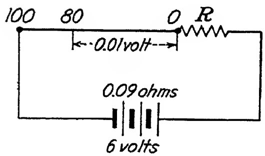


FIG. 163.—A potentiometer. The difference of potential across 80 cm. of wire is 0.01 volt. Illustrative Problem 6.

$$I = \frac{E}{R} = \frac{0.01 \text{ volt}}{1.76 \text{ ohms}} = 5.68 \times 10^{-3} \text{ amp.}$$

The total resistance of the circuit is therefore

$$R = \frac{E}{I} = \frac{6 \text{ volts}}{5.68 \times 10^{-3} \text{ amp.}} = 1{,}056 \text{ ohms.}$$

This is the sum of the resistances of each part of the circuit since they are all connected in series. 1,056 ohms = 0.09 ohm + R + 2.20 ohms. R = 1,056 ohms − 2.29 ohms = 1,054 ohms. Since the 1,056 ohms contain only four significant figures, only four figures of the result are significant.

7. If a coil wound with copper wire has a resistance of 100 ohms at 20°C., what will be its resistance at 45°C.?

The resistance at a temperature t°C. above 20°C. is given in terms of the resistance at 20°C. by the equation

$$R_t = R_{20}(1 + \theta t)$$

where $\theta = 0.0043 \frac{1}{°\text{C.}}$ is the temperature coefficient of resistance given here for copper at 18°C. from Table 15.

$$R_{45} = 100 \text{ ohms}\left[1 + 0.0043\frac{1}{°\text{C.}}(45°\text{C.} - 20°\text{C.})\right]$$
$$= 100 \text{ ohms } [1 + 0.1075] = 111 \text{ ohms.}$$

The change in the temperature coefficient of resistance between 18 and 20°C. is ignored.

8. What is the resistance of a coil that will heat 1 l. of water from 20 to 60°C. in 5 min. with 80 per cent efficiency when connected to a 115-volt line?

The energy required to heat the water is

$$H = mst$$

where m = 1 l. × 1,000 g./l. = the mass of the water.

s = 1 cal./g.-deg. = the number of calories to raise 1 g. of water 1°C.

t = 60°C. − 20°C. = the increase in temperature.

H = 1,000 g. × 1 cal./g. °C. × 40°C. = 40,000 cal.

= 40,000 cal. × 4.185 joules/cal. = 167,400 joules.

Since the heating coil is only 80 per cent efficient, the energy to be supplied by the electric current is 167,400/0.8 joules = 209,200 joules. This energy is supplied by the electric current in 5 min. = 300 sec. The rate of doing this work is, therefore, 209,200 joules/300 sec. = 697 watts. The power supplied by an electric current is $P = EI$, *cf.* Par. 20, or if Ohm's law is used, $P = E \times E/R = E^2/R$

$$P = 697 \text{ watts} = \frac{(115 \text{ volts})^2}{R}$$

$$R = \frac{(115 \text{ volts})^2}{697 \text{ watts}} = 19.0 \text{ ohms.}$$

9. A trolley line has a resistance of 0.5 ohm/mi. What is the drop in potential in the line if a car 3 mi. from the power house is using 60 amp.? What is the power lost in the line?

The total resistance of the trolley wire is 3 mi. × 0.5 ohm/mi. = 1.5 ohms. The resistance of the return path through the track may be neglected. The drop in potential in the line is, therefore,

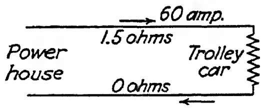


FIG. 164.—Resistance and current in a trolley line.

$$E = IR = 60 \text{ amp.} \times 1.5 \text{ ohms} = 90 \text{ volts.}$$

The power lost in the line is as in Prob. 8

$$P = EI = 90 \text{ volts} \times 60 \text{ amp.}$$
$$= 5{,}400 \text{ watts} = 5.4 \text{ kw.}$$

PROBLEMS

1. An electric lamp requires 115 volts. If it takes 1 amp., what is its resistance?

2. What is the relation between the resistance of four identical copper wires connected in parallel and that of one of the wires alone?

3. How many resistances may be made out of separate resistances of 2, 3, and 4 ohms?

4. What two resistances combined singly, in series and in parallel, will give resistances of 4, 5, 20, and 25 ohms?

5. A current of 50 amp. divides between three resistances of 3, 4, and 5 ohms in parallel. Find the current and difference of potential in each resistance.

6. A battery consists of six similar cells each of 2 volts e.m.f. The cells are connected in two parallel groups each consisting of three cells in series. What is the e.m.f. of the battery?

7. Two rheostats of 10- and 20-ohms resistance, respectively, are connected in series to a constant difference of potential of 120 volts. Find the current in the 10-ohm rheostat.

8. A storage battery consists of three cells each of 2 volts e.m.f. and internal resistance 0.05 ohm, all connected in series to a resistance coil. The current in the coil is 300 milliamp. Find the resistance of the coil.

9. A 6-volt storage battery has an internal resistance of 0.3 ohm. A lamp of 2-ohms resistance is connected (*a*) in series with a 2- and 4-ohm resistance and the storage battery, (*b*) in series with a 4-ohm resistance but with the 2-ohm resistance in parallel with the lamp. Find the ratio of the currents in the lamp in the two cases.

10. A cell of 2 volts e.m.f. and 0.5-ohm internal resistance is connected in series with a variable external resistance. Plot the terminal voltage of the cell as the external resistance varies from infinity to 0.

11. The generator of a car charges the storage battery at the rate of 20 amp. If the e.m.f. of the battery is 6 volts and its internal resistance 0.02 ohm, what difference of potential is required at the terminals of the generator?

12. A storage battery of three cells in series, each of e.m.f. 2 volts and internal resistance 0.01 ohm lights the two headlights of an automobile in parallel. If the car ammeter in series with the battery reads 8 amp. discharge, find the resistance of each lamp and the difference of potential between the terminals of each lamp.

13. If the lights in Prob. 12 are dimmed by putting a resistance in series with the battery, how much resistance would be required to cut the current in each lamp in half?

14. The current in a series circuit is 6 amp. When a resistance of 3 ohms is added, the current is 2 amp. What was the resistance of the original circuit?

15. What external resistance when connected across the terminals of a cell having an internal resistance of 1 ohm will make the potential difference between the terminals of the cell 0.6 of its e.m.f.?

16. A dry cell of 1.40 volts e.m.f. and 0.15-ohm internal resistance, a storage cell of 2.00 volts e.m.f. and 0.05-ohm internal resistance, and a coil of 2.8-ohms resistance are all connected in series with the positive terminal of the storage cell connected to the positive terminal of the dry cell. Find the current in the coil.

17. At the terminals of a certain generator in a power house, a voltmeter reads 130 volts when the generator is supplying a constant current of 40 amp. to a factory 3 mi. away. One mile of single wire such as is used in the two-wire line has a resistance of 0.05 ohm. What is difference of potential at the factory?

18. Three cells are connected as shown in Fig. 165. (*a*) has an e.m.f. of 2 volts and internal resistance of 0.3 ohm, (*b*) 1.5 volts and 0.2 ohm, and (*c*) 1 volt and 0.1 ohm. Find the current in each branch of the circuit.

19. A voltmeter with a resistance of 1,000 ohms reads 1 volt/mm. How can it be made to read 10 volts/mm.?

20. A galvanometer with a resistance of 20 ohms is deflected 1 mm. by a current of 1 milliamp. What resistance is required and how should it be connected to make a voltmeter reading 1 volt/mm.? What resistance is required and how should it be connected to make an ammeter reading 1 amp./mm.?

21. A voltmeter reads 25 volts full-scale deflection. The coil has a resistance of 500 ohms, and the series resistance is 2,000 ohms. What resistance must be shunted across the coil alone in order that it read 25 amp. full-scale deflection?

22. What fraction of the total current will flow through a galvanometer having a resistance of 10 ohms if shunted by a wire of 0.1-ohm resistance?

23. An ammeter gives its full-scale reading of 2 amp. with a potential difference of 0.01 volt across its terminals. If it is shunted with a resistance of 0.005 ohm, what line current will cause full-scale reading?

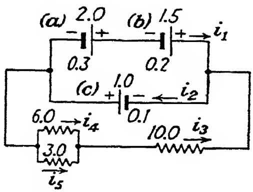


FIG. 165.—Three cells in a network. The e.m.f. of each cell in volts is given above it, the resistance in ohms below. The resistances in the circuit are in ohms. The direction of each current is marked by an arrow. If the current comes out negative, it flows in a direction opposite to that shown by the arrow. Prob. 18.

24. Suppose that the spring of a moving-coil galvanometer has a restoring torque on the coil proportional to the deflection, *i.e.*, $L = k\theta$. The coil has an area of 4 sq. cm., has 100 turns, a resistance of 10 ohms, and hangs in a radial magnetic field of 10^4 amp. turns/m. If a potential drop of 0.01 volt across the terminals gives a deflection of 45°, what is the value of k? How could this instrument be turned into a voltmeter giving a deflection of 45° for 150 volts?

25. A resistance of 2 ohms is connected in series with a 6-volt storage battery, with an internal resistance of 0.06 ohm. An ammeter with a resistance of 0.01 ohm and a voltmeter with a resistance of 500 ohms are used separately to measure the current and difference of potential of the 2-ohm resistance. How much error is made in each measurement and in the calculated resistance?

26. In the Wheatstone bridge of Fig. 158 the ratio of R_1/R_2 is 10. If R_3 is 42.6, what is the value of R_x? How much error in R_x would an error of 0.2 ohm in R_3 make?

27. In the Wheatstone bridge in Fig. 158, $R_1 = 40$ ohms, $R_2 = 10$ ohms, and $R_3 = 30$ ohms. What is R_x when no current flows through G? If current is supplied to the bridge by a 6-volt storage battery of 0.2-ohm internal resistance, how much power is dissipated in the bridge when it is balanced?

28. A constant current of 3 amp. flows in a long uniform wire. The difference of potential between two points on the wire 1 m. apart is 200 millivolts. Find the resistance of 10 m. of the wire.

29. A uniform straight wire 1 m. long is connected in series with a storage cell of 2 volts e.m.f. and 0.05-ohm internal resistance. The connecting wires have a resistance of 2.95 ohms and the current in the circuit is 0.2 amp.

How far apart are two points on the wire between which the difference of potential is 0.5 volt?

30. Suppose that in the potentiometer of Fig. 157 AB is a straight wire of specific resistance 14×10^{-8} ohm-m., diameter 0.6 mm., and length 1 m. If the galvanometer reads zero when P is 20 cm. from A and the current read on an ammeter in series with the battery is 0.25 amp., what is the value of the unknown potential?

31. A potentiometer consists of a slide wire on a meter bar. If there are 5 volts between the ends of the wire and an unknown potential requires 80 cm. of wire to balance it, what is the value of the unknown potential?

32. What is the resistance of a cylindrical copper wire 1 mm. in diameter and 1 m. long?

33. Two uniform conductors are made of the same material and have the same length. One, A, is a solid wire of 1 mm. diameter. The other, B, is a hollow tube with an outside diameter of 2 mm. and an inside diameter of 1 mm. Find the ratio of the resistance of A to the resistance of B.

34. The terminals of a voltmeter are successively connected to adjacent pairs of points A, B, and C, 10 cm. apart on a long straight nichrome wire carrying a current with the following consistently reproducible readings: $V_{AB} = 0.15$ volt, $V_{BC} = 0.20$ volt. If the current in the wire is 5 amp., find the cross section of each part of the wire.

35. Four identical copper wires are connected in series. A single wire of the same length is to have the same resistance as the four in series. What is the relation of its cross section to that of one of the four wires?

36. An iron wire and a copper wire of the same length and diameter are connected in series to a storage battery of 6 volts e.m.f. and 0.2-ohm internal resistance. The current is 0.5 amp. What is the resistance of each wire?

37. The resistance of a copper wire is 273 ohms at 0°C. What is the resistance of the wire at 100°C.?

38. Find the heat developed during 40 sec. by a 25-watt lamp.

39. Find the current in a heating coil on a 115-volt line necessary to develop 50,000 cal. in 10 min.

40. At 3 cts./kw.-hr., what is the cost of heating 1 cu. m. of water from 15 to 85°C. by electricity?

41. A current of 3 amp. at a difference of potential of 115 volts heats 1 kg. of water 24.5°C. in 5 min. by means of a heating coil immersed in the water. What is the mechanical equivalent of heat?

42. To prevent an undesirable temperature rise, a magnet is wound with copper tubing instead of copper wire. Through the tubing, a stream of water flows to carry away the heat. If the coils take 100 amp. from a 115-volt line, what flow of water would absorb all the heat with a 5°C. rise in temperature?

43. An electric light bulb operates at its normal rating when receiving 75 watts from a 115-volt line. What resistance must be connected in series with it in order that it may be operated at its normal rating from a 230-volt line?

44. Two 10-kw. 115-volt heating coils are connected in parallel to a 115-volt line. What is the ratio of the power consumed to that for the same two coils in series?

45. A storage battery supplies 10 watts to a circuit of 3.6 ohms resistance. What is the e.m.f. of the battery?

46. A certain 6-volt storage battery is capable of delivering 200 amp. hr. without appreciable drop in e.m.f. How much chemical energy is stored in the battery?

47. The potential difference at the power house is 120 volts and at the other end of the line 115 volts. The resistance of the line is 0.05 ohm. What power is lost in the line?

48. What diameter of copper transmission line wire must be used if 2,000 kw. of power are to be transmitted a distance of 100 km. with a loss of 5 per cent in transmission? Assume the potential at the generator to be (*a*) 1,000 volts and (*b*) 10^5 volts.

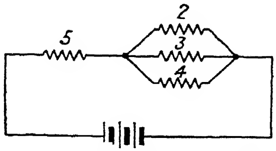


FIG. 166.—A storage battery connected to a resistance circuit. The resistances are in ohms. Prob. 52.

49. What power is required to light 10 lamps connected in parallel to a 115-volt circuit when each lamp takes 0.5 amp.?

50. Two coils A and B have resistances such that $R_A = 2R_B$. Find the ratio of the potential difference applied to A to that applied to B in order that the power dissipated in B may be twice that dissipated in A.

51. Two coils A and B of resistance R_A and R_B, respectively, are connected in series to a source of constant potential. The current in A is I_A, and that in B is I_B. Find the ratio of the power expended in A to the power expended in B.

52. A storage battery of three cells each of e.m.f. 2 volts and internal resistance 0.1 ohm is connected in the circuit as shown in Fig. 166. Find (*a*) the current in the battery, (*b*) the rate of dissipation of energy in the 3-ohm resistance.

53. A 100-watt incandescent lamp operates on 115 volts. What is the resistance of the lamp? How much energy does it consume in 1 hr.?

CHAPTER XVIII

ROTATIONAL MOTION

1. In recent chapters, we have frequently spoken of the forces on electric charges and the motion caused by these forces but have not had occasion to investigate such motions in detail. The further development of the subject of electricity makes it necessary that we do so. Furthermore, we find that the motions that we need to study are of a somewhat different kind from the simple linear motions that we studied early in the book. In this chapter, we shall supplement our earlier treatment of uniformly accelerated linear motion with a discussion of circular motion. In a later chapter, we shall take up a case of varying acceleration of a particular type known as simple harmonic motion.

Most General Type of Motion of a Rigid Body

2. In Chap. III, Par. 8, it was explained that the position of a point in three-dimensional space was completely specified by three numbers which we called the coordinates of that point. The motion of such a point in space could then be described in terms of the changes of these three coordinates. Of the several examples which were given, that of a car driving from the Capitol to the Lincoln Memorial may be recalled. In talking about the motion of this car, it was treated as if it were a point, as if it were of no interest whether the car was right side up or upside down, headed down the street or across it. Obviously, this is an over-simplified description of the position of the car. We might often want to know the orientation of the car as well as the position of its center. We shall continue to make the problem somewhat simpler than it really is by considering the car as a "rigid body," *i.e.*, one in which the distance from any one part of the body to any other part of the body remains constant. We know that the position of the center of such a body is determined by three numbers. The question is, then, how many more numbers are

needed to specify completely the orientation of such a body? The answer is three. That this is the answer can be shown in several ways. We can think of the possible changes in position of a body, one point of which is fixed, as made up of rotations about three mutually perpendicular axes in space, and we can specify the amount of these rotations by three angles. A little thought supplemented by an experiment with a book, a cake of soap, or any other rigid body will show that all possible orientations of a rigid body can be obtained in this way. A more analytical way of reaching the answer "three" is to start from the realization that the position of the body is determined if the positions of three points in it are known. These positions are specified by nine coordinates, but these nine coordinates are not entirely independent since the three separations of the points must remain unchanged. Therefore, the nine coordinates are related by three equations which reduce the number of independent unknowns to six. But three of these are necessary to fix the position of one of the points in space so that only three more, evidently, are necessary to fix the orientation of the body.

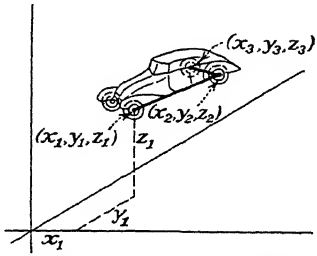
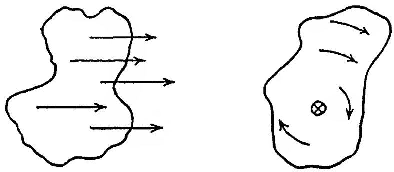


FIG. 167.—Schematic drawing of the specification of position of a rigid body in space by the coordinates of any three points in the body.

3. Going back to the specific example of the car on a road, suppose it has skidded into the ditch and we want to know its position exactly. Three coordinates will give the exact location in space of some one point, say the left front hub cap; the amount that the car has twisted off its course in a horizontal plane can be specified by giving the angle that a line through the two left hub caps makes with a vertical north-south plane; the amount that the front of the car has fallen or risen can be specified by the angle this same line makes with the horizontal; finally, the sideways tilt of the car can be specified by the angle that a line through the two back hub caps makes with the horizontal. Thus the position and orientation have been completely specified by six quantities exemplifying the general result that holds for any rigid body.

4. We might point out in passing that if the body is not rigid, still further coordinates have to be introduced to specify the relative positions of its parts and still further types of motion are possible. The different types of motion possible for a body are called degrees of freedom. Thus a rigid body moving unconstrained through space has six degrees of freedom. A particle so small that it can be considered as a point has three degrees of freedom. A car moving on a straight track has one degree of freedom. A ball on a smooth table has five degrees of freedom, and so on.

5. If a body is moving in such a way that the paths of all parts of the body are parallel to each other as in Fig. 168(a), it is



(a)-Pure translation *(b)*-Pure rotation
FIG. 168.—Two types of motion.

said to have a motion of pure translation. If one point of the body is fixed and the others revolve around it, the body is said to have a motion of pure rotation. It is possible to show that all motions of rigid bodies can be considered as made up of a pure translation of the center of mass of the body combined with a pure rotation of the body around that point.

6. The consideration of such a general motion of a rigid body where all six degrees of freedom come into play is far beyond the scope of this book. We have already considered translational motion in one or two dimensions, *i.e.*, with one or two degrees of freedom, and even, in the kinetic theory of gases, considered three-dimensional motion in a simplified way. To this treatment, we are now going to add the consideration of pure rotational motion about a single axis, *i.e.*, with a single degree of freedom. This kind of motion is, of course, continually met in machinery of all sorts where wheels are moving around fixed axes. This kind of motion is called circular motion.

Circular Motion

7. The first thing to do in studying any kind of motion is to choose coordinates that will be suitable. In the case of a wheel moving around a fixed axis, we could specify its position by giving ordinary x,y-coordinates of some particular point on the wheel referred to two fixed coordinate axes in the plane of the wheel; but it is much simpler to specify its position by giving the angle which some particular radius of the wheel, a certain spoke, for example, makes with a stationary line along a diameter of the wheel. Thus, in Fig. 169, the angle between the line OB on the rotating wheel and the fixed line OA increases as the wheel turns, and its value at any time gives the position of the wheel. It is the rotational coordinate of the wheel. It is customary to measure it as positive in a counterclockwise direction. Also, it is usually more convenient to measure it in radians than in degrees. (It may be recalled that the ratio of the length of the arc to the length of the radius gives the angle in radians, and since this is a ratio of two quantities of the same kind, it is itself a pure number. Thus, in Fig. 169,

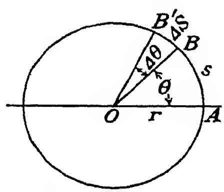


FIG. 169.—A wheel rotating about fixed axis. θ measures the angular and s the linear distance of rotation.

$$\theta = \frac{AB}{OA} = \frac{s}{r}\text{ rad.} \qquad \text{or} \qquad s = r\theta.)$$

Angular Displacement, Velocity, and Acceleration

8. In translational motion, a change in the position of a body was called a displacement. Similarly, the change in the position of a rotating body may be called an angular displacement. Such a displacement is usually denoted by θ. The rate of change of angular position, or the angular displacement per unit time, is the angular velocity of a rotating body. It is usually designated by the Greek letter omega ω and is defined mathematically as

$$\omega = \frac{\Delta\theta}{\Delta t}$$

where $\Delta\theta$ is the change in θ during the time Δt. There is a real difference between rotational and translational motion which

affects the units in which angular velocities are expressed. In rotational motion, the moving body comes back every so often to the position from which it started, *i.e.*, θ varies only between 0 and 2π. When this happens, the body is said to have made one revolution. Angular velocities expressed in revolutions per second or per minute are sometimes more useful than angular velocities expressed in radians or degrees per second. Since one revolution corresponds to an angular displacement of 2π rad., $\omega = 2\pi n$ rad./sec. if n is the number of revolutions per second.

9. Angular velocity is a vector quantity represented by a vector that lies along the axis of rotation in the direction in which a right-hand screw would be driven by the rotation. The magnitude of the vector is proportional to the angular speed.

10. Angular acceleration is the rate of change of angular velocity. It is usually designated by the Greek letter alpha α and is defined mathematically as

$$\alpha = \frac{\Delta\omega}{\Delta t}$$

where $\Delta\omega$ is the change of angular velocity in the time Δt. In terms of the calculus, the angular velocity is

$$\omega = \frac{d\theta}{dt}$$

and the angular acceleration is

$$\alpha = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2}.$$

The angular acceleration is expressed in radians per second per second or sometimes in revolutions per second per second. The m.k.s. units of angular motion, *i.e.*, the units that are to be used in the various equations which we shall derive, are the radian, the radian per second, and the radian per second per second.

Equations of Angular or Circular Motion

11. By a process of reasoning exactly analogous to that used for uniformly accelerated linear motion (Chap. III, Par. 24), we can write down corresponding equations for uniformly accelerated rotational motion. For convenience, we shall repeat

the equations for linear motion and include the relation between n and ω mentioned above. We have also added a relation defining the period T, which is the time required for one complete revolution. It is a very useful quantity but obviously has significance only when ω is constant.

Linear Motion		Angular Motion	
$v_t = v_0 + at$	(1)	$\omega_t = \omega_0 + \alpha t$	(4)
$s = v_0 t + \frac{1}{2}at^2$	(2)	$\theta = \omega_0 t + \frac{1}{2}\alpha t^2$	(5)
$2as = v_t^2 - v_0^2$	(3)	$2\alpha\theta = \omega_t^2 - \omega_0^2$	(6)
		$\omega = 2\pi n, \qquad T = \dfrac{1}{n}$	(7)

The student is urged to verify these relations for angular motion by going through the derivations himself.

Relations between Linear and Angular Quantities

12. Instead of considering the motion of a rotating body as a whole, we now want to consider the motion of a particular point on it. Suppose we take the point P a distance r from the axis of a wheel rotating with an angular velocity ω. As shown in Fig. 170, after an instant of time Δt, P will have moved to P' and the radius OP will have turned through the angle $POP' = \Delta\theta$. Then by definition, the linear velocity of P is $v = PP'/\Delta t$ and the angular velocity

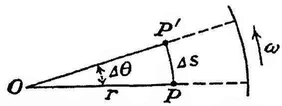


FIG. 170.—Relations between linear and angular motion.

$$\frac{\angle POP'}{\Delta t} \quad \frac{\Delta\theta}{\Delta t}.$$

But if the angle is small, the chord PP' equals the arc $r\Delta\theta$. Therefore,

$$v = \frac{r\Delta\theta}{\Delta t} = \omega r$$

which is the general relation between the linear velocity of a point distant r from the axis of rotation and the angular velocity ω about that axis.

13. If the linear velocity is changing from a value v_0, at a time t_0, to a value v_1, at a time t_1, then, by definition, the linear acceleration is given by

$$a = \frac{v_1 - v_0}{t_1 - t_0} = \frac{\omega_1 r - \omega_0 r}{t_1 - t_0} = r\frac{\Delta\omega}{\Delta t}$$

or

$$a = r\alpha.$$

This is the general relation between the angular and linear acceleration of a point on a rotating body. These relations between linear and angular motions of the same body are very useful and may be repeated in a group as follows:

$$\begin{aligned} s &= r\theta \\ v &= r\omega \qquad (8) \\ a &= r\alpha. \end{aligned}$$

Centripetal and Centrifugal Force

14. If a weight on the end of a string is swung around in a circle, it pulls on the string because of the "centrifugal force."

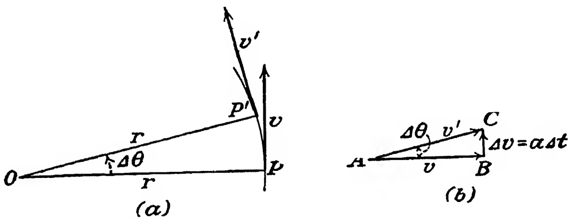


FIG. 171.—Velocity and acceleration in pure rotation. (*a*) The motion. (*b*) Vector diagram of velocities.

What is "centrifugal force"? To answer this question, the conditions satisfied by a body moving in a circular path must be investigated. Suppose the weight is moving in a circular path of radius r, with a linear velocity v of constant magnitude. The velocity will always be tangential to the path, and since the path is curved, the direction of the velocity must be continually changing. This means there is an acceleration, and our problem is to investigate its direction and magnitude. Suppose that, as shown in Fig. 171(*a*), the body is at the point P with the velocity v and then after a time Δt it has moved to P', a distance $\Delta s = v\Delta t$, and its velocity has changed to v' where v' is equal in magnitude but different in direction from v. Since v is perpendicular to OP and v' is perpendicular to OP', the angle between v and v' must

be the $\angle POP' = \Delta\theta$. Now draw a vector diagram as in Fig. 171(b) where the two velocities are represented by the vectors AB and AC. Then the vector BC must be the change in the velocity during the time interval Δt since the vector BC added vectorially to the old velocity AB gives the new velocity AC. But by definition, if a is the acceleration, $a\Delta t$ is the change in the velocity in the time Δt; therefore, $BC = \Delta v = a\Delta t$. Now consider the geometry of Figs. 171(a) and 171(b). The angle $\Delta\theta$ is the same in each case and in each case can be measured by the ratio of the arc (or chord if the angle is very small) to the radius. This gives us

$$\text{From } (a)\ \Delta\theta = \frac{\Delta s}{r} = \frac{v\Delta t}{r} \tag{9}$$

$$\text{From } (b)\ \Delta\theta = \frac{\Delta v}{v} = \frac{a\Delta t}{v}.$$

Combining these equations, we have

$$a\frac{\Delta t}{v} = v\frac{\Delta t}{r} \tag{10}$$

or

$$a = \frac{v^2}{r}.$$

In terms of angular velocity, this becomes $a = \omega^2 r$. It is clear from the figure that the direction of the acceleration is perpendicular to the linear velocity and is therefore directed inward along the radius of the circle in which the body is moving. The magnitude of the acceleration is proportional to the square of the linear velocity and inversely proportional to the radius.

15. Where there is an acceleration, there must be a force. To put it another way, according to Newton's first law, a body will move in a straight line unless there is a force acting on it, so that if a body is moving in a circular path a force must be acting on it. According to Newton's second law, $F = ma$. We see, therefore, that *a force of $mv^2/r = m\omega^2 r$ newtons perpendicular to the instantaneous direction of motion of a body of mass m kilograms is* necessary and sufficient to *cause that body to move in a circle of radius r meters. Such a force is called a centripetal force.* To every

action, there is an equal and opposite reaction, and the reaction against the centripetal force is called the centrifugal force. It is the force of reaction of the body against restriction to a circular path. Obviously, the centrifugal force is also of magnitude $mv^2/r = m\omega^2 r$ and is directed outward along the radius of curvature of the path of the body.

Angular Acceleration and Torque. Moment of Inertia

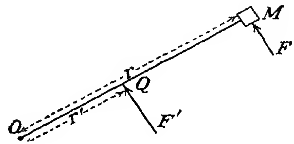
16. In the last section, we found it convenient to discuss rotational motion in terms of the instantaneous linear velocity and acceleration of a particle moving in a circular path. In this section, we shall use a similar procedure to derive the relation between rotational acceleration and rotational force. We know that the effect of a force in setting up a rotation depends not only on the magnitude of the force but on its line of application. If a force is applied to the center of a wheel mounted on a fixed axis, it produces no effect; if applied at the rim of the wheel but pointing toward the center, it produces no effect; but if it is applied tangential to the rim, it produces a maximum effect. In short, it is the moment of the force or torque that determines the rotational action. This differs from the relation in translational motion, the relation expressed by the equation $F = ma$. Does the inertial reaction of the wheel also differ? It does. The rotational inertia of a body depends not only on its mass but on the distribution of that mass. It is harder to get a wheel turning if most of its mass is in the outer rim than if most of its mass is near the axis. Qualitatively, then, we can say that the angular acceleration of a rotating body depends on the sum of the moments of force acting on it, on its mass, and on its distribution of mass. Just what this dependence is remains to be considered.

17. This can be done most easily by returning to the case of the single weight rotating at the end of a string or, better, at the end of a rigid weightless rod (Fig. 172). Suppose that in addition to the centripetal force exerted on the weight by this rod there is another force F acting on the weight perpendicular to the rod. Then the linear velocity v will change in both magnitude and direction. The rate of change of its magnitude is what now concerns us. It will be $a = F/m$. The moment of force of F about O is Fr. The same turning effect can be produced by any other force F' acting at such a point Q on the rigid rod that its moment of

force $F'r'$ about O is equal to Fr. Call this moment of force $Fr = L$. Then the linear acceleration

$$a = F/m = Fr/mr = L/mr.$$

L is the turning moment or torque of the force setting m in rotation. The inertial reaction against this torque is mar. Suppose now that we have a number of masses $m_1, m_2, \ldots$, etc., at distances $r_1, r_2, \ldots$ along the rigid rod from O and a number of forces acting having moments around O of $L_1, L_2, L_3, \ldots$ Then the sum of the inertial reactions against rotational acceleration must be equal to the sum of the torques tending to produce it, *i.e.*,



torque moment of a force. The force F acts at a perpendicular distance r from O and the force F' at a distance r'.

$$L_1 + L_2 + L_3 + \qquad = m_1r_1a_1 + m_2r_2a_2 + \tag{11}$$

Call the total torque L, and substitute for the various linear acceleration a_1, a_2, etc., their equivalents αr_1, αr_2, αr_3, etc., where α is the angular acceleration which is the same for all the masses. This gives

$$L = (m_1r_1^2 + m_2r_2^2 + m_3r_3^2 + \qquad)\alpha. \tag{12}$$

This expression was derived in terms of masses distributed along a weightless rod, but the derivation is still valid if we consider the masses increased in number until they form a continuous solid rod or distributed not only along one line radiating from the axis but along any number of such lines. By such processes, any distribution of mass can be built up and any rigid body represented. The quantity in parentheses in Eq. (12) is called the moment of inertia of a body. It is usually designated by I. Equation (12) can, therefore, be written

$$L = I\alpha \tag{13}$$

and this is the relation between torque and angular acceleration that corresponds to the equation $F = ma$ in translational motion.

Determination of Moments of Inertia

18. It is evident that the moment of inertia of a body depends both on the geometrical distribution of mass in the body and on the location of the axis of rotation. It is obtained by a process of summation as is indicated by Eq. (12). In a few cases, this can be done by elementary methods. For example, the moment of inertia of a flywheel most of whose mass is concentrated in the rim is obviously approximately equal to the mass of the flywheel times the square of the radius of the rim if the axis of rotation is the axle of the wheel. But most calculations of moments of inertia, if they can be carried out at all, are done by the use of integral calculus. We shall not attempt to give an example of the process but shall quote some results in Table 16.

TABLE 16.—MOMENTS OF INERTIA

Body	Axis	Moment of inertia
Uniform thin rod..........	Normal to length at one end	$\frac{ML^2}{3}$
Uniform thin rod..........	Normal to length at center	$\frac{ML^2}{12}$
Thin circular sheet of radius R.....................	Normal to the plate through the center	$\frac{MR^2}{2}$
Thin circular sheet of radius R.....................	Along any diameter	$\frac{MR^2}{4}$
Rectangular parallelopiped, edges a, b, c.............	Through center perpendicular to face ab	$\frac{M(a^2 + b^2)}{12}$
Sphere of radius R.........	Any diameter	$\frac{2MR^2}{5}$
Right circular cylinder of radius R and length L....	Longitudinal axis of cylinder	$\frac{MR^2}{2}$
Right circular cylinder of radius R and length L....	Through center perpendicular to axis of cylinder	$M\left(\frac{R^2}{4} + \frac{L^2}{12}\right)$

Kinetic Energy of a Rotating Body

19. At any instant, the kinetic energy of any moving body will be given by

$$\text{K.E.} = \tfrac{1}{2}(m_1v_1^2 + m_2v_2^2 + m_3v_3^2 + \cdots) \tag{14}$$

where m_1, m_2, m_3, etc., are the small bits of mass that compose the body and v_1, v_2, v_3, etc., are the instantaneous linear velocities of these masses. If the body is rotating about a fixed axis with an angular velocity ω, $v_1 = \omega r_1$, $v_2 = \omega r_2$, etc., where r_1, r_2, . . . are the distances of the masses from the axis, so that the foregoing expression becomes

$$\text{K.E.} = \tfrac{1}{2}(m_1\omega^2 r_1^2 + m_2\omega^2 r_2^2 + \cdots) = \tfrac{1}{2}\omega^2 \Sigma m r^2$$

but $\Sigma mr^2 = I$, the moment of inertia. Therefore

$$\text{K.E.} = \tfrac{1}{2}I\omega^2. \tag{15}$$

If a rotating body of moment of inertia I is accelerated from an angular velocity 0 to an angular velocity ω in rotating through an angle θ, it acquires a kinetic energy $\frac{1}{2}I\omega^2$, and this must equal the work that has been done on the body in moving it through the angle θ. But we have from Eq. (6) that

$$2\alpha\theta = \omega^2.$$

Therefore, the work done is

$$I\alpha\theta \tag{16}$$

but $I\alpha = L$ the torque; therefore, the work done by a torque L turning a body through an angle θ is

$$\text{Work} = L\theta. \tag{17}$$

Angular Momentum

20. Since I and ω play the roles in rotational motion corresponding to m and v in translational motion, it is natural to *define* their product *$I\omega$ as angular momentum.* Angular momentum is a vector quantity whose direction is determined by the direction of the axis of rotation in the same way that the direction of ω is determined. The relation of angular momentum to linear momentum can be understood best by returning to the simplified example of rotational motion illustrated in Fig. 172. Evidently, the rotating body consisting of the mass m at the end of the weightless bar OM has a moment of inertia mr^2. By the definition just given, its angular momentum about O must therefore be ωmr^2; but ωr equals v, the linear velocity of m; therefore, the angular momentum $I\omega$ equals mvr, the linear momentum times

the distance from the axis of rotation. For this reason, the angular momentum is sometimes called the *moment of momentum*. Extending the preceding considerations to every particle of a rigid body rotating about an axis, we obtain the general relation

$$\omega I = \omega \Sigma m_i r_i^2 = \Sigma m_i \omega r_i^2 = \Sigma m_i v_i r_i.$$

21. It is natural to expect from the consistency of our analogies between linear and rotational motion that *the rate of change of the angular momentum of a rotating body* should be *proportional to the torque acting on it* just as the rate of change of the linear momentum of a body is proportional to the force acting. That this is indeed true is made clear by returning once more to the simple case of Fig. 172, the body of mass m rotating at a distance r about an axis at O. If the force acting is F, the rate of change of linear momentum is $\Delta(mv)/\Delta t = F$. Multiplying by r, we get $\Delta(mrv)/\Delta t = Fr$. Substituting ωr for v gives $\Delta(mr^2\omega)/\Delta t = Fr$, but mr^2 is the moment of inertia and Fr is the torque; therefore,

$$\frac{\Delta(I\omega)}{\Delta t} = L.$$

This is essentially a repetition of the argument of Par. 17.

Conservation of Angular Momentum

22. The italicized statement in the last paragraph is simply Newton's second law of motion formulated in a way directly applicable to rotational motion. In Par. 24, Chap. IV, we deduced from Newton's second law the principle of conservation of linear momentum. Similarly, it is evident from the form of the law appropriate to rotational motion that if the value of the resultant torque on a body is zero its rate of change of angular momentum is zero. This gives the principle of conservation of angular momentum as follows:

The total angular momentum of a system of bodies on which no external torque is acting remains the same. It is not changed by any reactions between bodies within the system.

The Gyroscope

23. Of the many examples of the conservation of angular momentum that might be described, probably the simple top or gyroscope is the most interesting. Perhaps the most useful

form of gyroscope for demonstration purposes is a bicycle wheel with a lead tire and with handles on the axle. When such a wheel is rotating at a high speed, it will resist any attempt to turn its axis of rotation almost as if it were alive. Since it has a large moment of inertia and a high angular velocity, any such change in the axis of rotation means a change from big angular momentum in one direction to a big angular momentum in another direction, *i.e.*, a big change in angular momentum; therefore, a large torque is necessary to produce this change in a short time. Besides this direct reaction to a change in the direction of the angular momentum, the conservation of angular momentum can be observed when the bicycle wheel is made part of a system on which no outside torques can act. This may be done by having the man holding the wheel step on to a platform that is free to move on ball bearings around a vertical axis. Suppose that the man steps on the platform holding the wheel so that its angular-momentum vector is pointed straight up (*i.e.*, the wheel is turning in a horizontal plane in a counterclockwise direction when viewed from above). The resultant angular momentum of the man and the wheel is then in that same direction. Now suppose the man turns the wheel over so that its angular momentum vector is pointed down. No external torque around a vertical axis can act through the pivot of the stand so that the resultant angular momentum around a vertical axis must remain unchanged. Consequently, some motion must appear that will give an angular-momentum vector upward equal to twice the downward momentum of the wheel. This does in fact occur, the whole system of man, bicycle wheel, and turntable rotating counterclockwise until friction in the bearings of the turntable brings them to a stop. The bicycle wheel can then be turned upside down, again producing clockwise rotation of the system, and so on, until the bicycle wheel slows down. Fortunately for the man performing the experiment, the combined moment of inertia of himself, the bicycle wheel, and the turntable is enough larger than that of the bicycle wheel alone so that the angular velocity of the whole system does not have to be comparable with that of the wheel to have their angular momenta the same order of magnitude.

24. Another striking property of rotating bodies that can be demonstrated with this same bicycle wheel is the phenomenon

known as precession. Suppose that the wheel is spinning with its axis horizontal and that it is supported only at one end of the axle. We know that it has a big angular momentum around a horizontal axis and that it will take a large torque to change this. Actually, the torque acting is the weight of the wheel times the distance from its center of mass to the point of support of the axle. This will tend to turn the wheel about a horizontal axis perpendicular to that around which it is spinning. But it may not be big enough to push the spin axis down very rapidly. What we actually observe is that there is

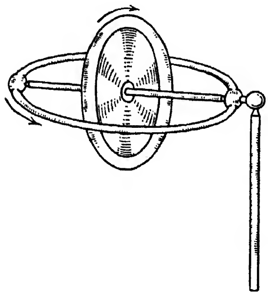


FIG. 173.—Gyroscopic top. The top spins about a horizontal axis. A torque due to the weight acts on the top about a horizontal axis at right angles to the first. Because of this torque the top precesses about a vertical

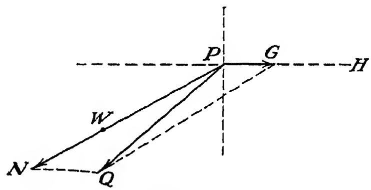


FIG. 174.—Angular velocity vectors in a gyroscope. *PN* represents the spin angular momentum, *PG* the angular momentum due to the external torque, and *PQ* the resultant angular momentum.

a hardly perceptible falling of this spin axis below the horizontal plane but that it does move in that plane. This motion, which may be understood more clearly from the diagram in Fig. 174, is called precession. The most satisfactory elementary explanation of precession is in terms of angular-momentum vectors. In Fig. 174, let *PN* represent the original angular momentum $I\omega$. Then the weight of the wheel acting at *W* distant *PW* from the support of the end of the axle introduces a torque around the horizontal axis *PH*, and this torque will produce an angular momentum represented by the vector *PG*. The resultant angular momentum of the wheel will be the vector sum of this new momentum and the previous spin *PN*. This resultant angular momentum is represented by the vector *PQ* and is evidently in the horizontal plane but at an angle to the original spin. This turning of the axis of resultant angular momentum continues as long as the

torque of the weight acts. Consequently, the spinning wheel continually precesses around its point of support.

Recapitulation

25. In Par. 11, we wrote down beside each other the so-called kinematical relations for translational and rotational motion. These are the equations that describe the motions without saying anything about their cause. In later paragraphs, we derived various relations between these motions and the forces and energies involved in them. It may be well to collect some of these relations and compare them in tabular form with the corresponding ones for translation.

Translation		Rotation	
Inertia	m	Moment of inertia	$I = \Sigma mr^2$
Force	$F = ma$	Torque	$L = I\alpha$
Work	Fs	Work	$L\theta$
K.E.	$\frac{1}{2}mv^2$	K.E.	$\frac{1}{2}I\omega^2$
Momentum	mv	Angular momentum	$I\omega$

(18)

SUMMARY

An extended rigid body requires six coordinates for the specification of its position. The most general motion of a rigid body consists of a translational motion of the center of mass plus a rotation of the body around an axis through the center of mass. In this chapter, we consider rotation about a fixed axis only. The only convenient coordinate to specify position in such motion is the angle between a line rotating with the body and a line fixed in space. The definitions of angular displacement θ, velocity ω, and acceleration α are exactly analogous to the corresponding definitions in linear motion. Equations relating them are analogous to those for linear motion also. The linear displacement, velocity, and acceleration for a point a distance r from the axis of rotation are given by the relations $s = r\theta$, $v = r\omega$, and $a = r\alpha$.

The centripetal force necessary to hold a mass m in a circular path of radius r is $mv^2/r = m\omega^2r$ directed toward the axis of rotation. The effect of inertia in resisting angular acceleration depends not only on mass but on distribution of mass. For a given body and axis of rotation, this inertial reaction is given by

$\Sigma mr^2 = I$ where m is the mass of a part of the body at a distance r from the axis and the summation is taken over the whole body. I is called the moment of inertia. Corresponding to the $F = ma$ equation in linear motion, we have $L = I\alpha$ for rotational motion where L is the applied torque or moment of force. A table of moments of inertia is given.

The kinetic energy of a rotating body is $\frac{1}{2}I\omega^2$. The angular momentum is $I\omega$. If no external torque is acting on a system of bodies, the resultant angular momentum remains constant. This principle has many applications of which the gyroscope is one of the most interesting.

The chapter ends with a table of corresponding quantities in translational and rotational motion.

ILLUSTRATIVE PROBLEMS

1. What is the linear velocity of a point 20 cm. from the axis of a wheel making 2.5 turns per second? Also, what is the velocity of a point 50 cm. from the axis? What is the angular velocity of each point?

In one turn, a point at a distance r from the axis of the wheel moves through an arc of length $2\pi r$. From the definition of a radian in Par. 7, this is equivalent to an angle of $2\pi r/r = 2\pi$ rad. The angle turned through is seen to be independent of the distance of the point from the axis of the wheel. From this definition of a radian, it also follows that a radian is a pure number, a quantity without the physical dimensions of mass, length, or time.

The wheel has an angular velocity of 2.5 turns per second. This is equivalent to 2.5 turns/sec. $\times$ 2π rad./turn = 15.71 rad./sec. This is the angular velocity of all points of the wheel and, hence, of the two points 20 and 50 cm. from the axis.

The linear velocity is given in Par. 12 as $v = \omega r$ where ω is the angular velocity and r the distance of the point from the axis. Thus,

$$v = 15.71\ \frac{\text{rad.}}{\text{sec.}} \times 20\ \text{cm.} = 314\ \text{cm./sec.}$$

and

$$15.71\ \frac{\text{rad.}}{\text{sec.}} \times 50\ \text{cm.} \quad 786\ \text{cm./sec.}$$

2. In an Atwood's machine (*cf.* Chap. IV, illustrative Prob. 4), two weights of 6 kg. each are connected by a cord and suspended from a pulley. The mass of the cord and the friction of the pulley may be neglected. Suppose that the pulley is a right circular cylinder with a mass of 4 kg., and assume that the cord does not slip on the pulley. If now a weight of 2 kg. is added to one side, find the downward acceleration of that side and the tension in the string.

In this problem there are three bodies to be accelerated. We can consider the forces on them separately and then consider the relations between them. Evidently the right-hand mass is going to fall, the cylinder to rotate clockwise, and the left-hand weight to rise. Call these directions positive and the reverse directions negative. Let m_1 be the mass on the right, m_2 the mass on the left and I the moment of inertia of the cylinder. Let r be the radius of the cylinder, a_1 the linear acceleration of m_1, a_2 the linear acceleration of m_2, and α the angular acceleration of the cylinder. Then for the linear motions in each case the total force must equal the mass times the acceleration, giving

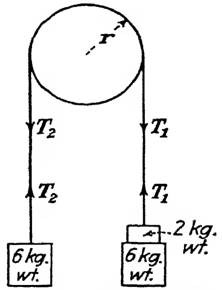


FIG. 175.—Atwood's machine. The tensions in the two strings are not the same since some tension must be used to accelerate the pulley.

for m_1 $\qquad m_1g - T_1 \qquad m_1a_1$
for m_2 $\qquad T_2 - m_2g \qquad m_2a_2.$

In the rotational motion the total torque equals the moment of inertia times the angular acceleration. Each tension acts with a moment arm r so that we have

for the cylinder $\qquad (T_1 - T_2)r = I\alpha.$

If the string does not stretch or slip, $a_1 = a_2 = \alpha r$. Dropping the subscripts on a, substituting a/r for α and dividing the equation for the cylinder by r we have

$$m_1g - T_1 = m_1a \tag{19}$$
$$T_2 - m_2g = m_2a \tag{20}$$
$$T_1 - T_2 = Ia/r^2 \tag{21}$$

When these three equations are added the tensions disappear, giving

$$m_1g - m_2g = m_1a + m_2a + Ia/r^2$$

$$\frac{(m_1 - m_2)g}{m_1 + m_2 + I/r^2} \tag{22}$$

Before substituting numbers in this result we must compute I/r^2. From Table 16 we know that $I = Mr^2/2$ where M is the mass of the cylinder. Therefore $I/r^2 = M/2 = 2$ kg. Substituting this and the other numbers given in the statement of the problem the acceleration of the system is given by

$$a = \frac{(8 \text{ kg.} - 6 \text{ kg.}) \times 9.8 \text{ m./sec.}^2}{8 \text{ kg.} + 6 \text{ kg.} + 2 \text{ kg.}}$$
$$= \frac{2 \text{ kg.} \times 9.8 \text{ m./sec.}^2}{16 \text{ kg.}}$$
$$= 1.225 \text{ m./sec.}^2$$

To obtain the tensions substitute in Eqs. (19) and (20)

$$\begin{aligned} T_1 &= m_1g - m_1a = 8 \text{ kg.} \times (9.8 \text{ m./sec.}^2 - 1.225 \text{ m./sec.}^2) \\ &= 8 \text{ kg.} \times 8.575 \text{ m./sec.}^2 = 68.6 \text{ newtons} \\ T_2 &= m_2a + m_2g = 6 \text{ kg.} \times (9.8 \text{ m./sec.}^2 + 1.225 \text{ m./sec.}^2) \\ &= 6 \text{ kg.} \times 11.025 \text{ m./sec.}^2 = 66.2 \text{ newtons} \end{aligned}$$

3. How far above the top of the circle must a car start in order to be able to loop the loop from the effect of the force of gravity alone? Neglect friction.

Let h be the initial height of the car above the circle and r the radius of the circle. When the car is at the top of the loop, it must just be in contact with the track. Therefore, the force of gravity down must be balanced by the centrifugal force up. Thus, from Par. 15,

$$mg = \cdot \tag{23}$$

The velocity at the top of the loop will be the same as that obtained by falling freely from a height h. The velocity is the same at the two points A and B, Fig. 176, as the velocity gained from B to the bottom of the loop is lost again from the bottom up to A. This follows directly from the conservation of energy. The same kinetic energy $\frac{1}{2}mv^2 = mg2r$ is lost in rising from the bottom to A as is gained in falling from B. Since the kinetic energies at A and B are equal, the velocities are equal. The velocity at A is, therefore, Eq. (14), page 65, given by the equation

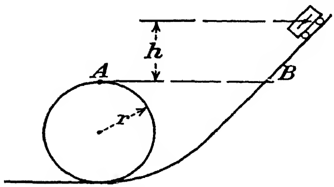


FIG. 176.—Loop the loop. The car starts at a height h above the top of a loop of radius r.

$$v^2 = 2gh.$$

Substituting this in Eq. (23), after canceling m, gives

$$g = \frac{2gh}{r} \qquad \text{so that} \qquad h = \frac{r}{2}.$$

PROBLEMS

1. A wheel spins with a constant angular velocity of 50 r.p.s. Through what angle does it revolve in 1 min.? What is the linear velocity of a point 10 cm. from the axis?

2. A wheel has a constant angular acceleration of 4 rad./sec.^{2} If it has an initial angular velocity of 10 rad./sec., how long will it take before its velocity is doubled? How long before it has made 10 rev.?

3. What acceleration is required to increase the angular velocity of a wheel from 10 to 30 r.p.s. in 5 rev.?

4. What is the angular velocity of a wheel of 50 cm. radius rolling along the ground with a velocity of 20 m./sec.?

5. If the moment of inertia of the coil in illustrative Problem 4, Chap. XVI, is 10^{-6} kg. m.^{2}, what is its angular acceleration?

6. Suppose that the turning disk in Barlow's wheel has a radius of R m. and a moment of inertia of I kg. m.^{2}, that the current flowing from the axis to the rim of the disk is i amp., and that in the outermost 2 cm. of the disk it passes through a uniform magnetic field of H amp. turns/m. What is the angular velocity of the disk t sec. after the current is turned on, and how many revolutions will have been made in that time?

7. A 5-kg. cylinder of 20 cm. radius is set rotating by wrapping a string around it and pulling with a force of 3-kg. weight on the string. How many revolutions will it make from rest in 5 sec.? What is the linear acceleration of the string?

8. A mass of 102 g. hangs from a string that is wound around the axle of a wheel. If the axle is 20 cm. in diameter and the wheel has a moment of inertia of 10^{-3} kg.-m.^{2}, what is its angular acceleration? If it starts from rest, how long will it take the weight to fall 50 cm.?

9. A 15-kg. sphere of 10 cm. radius is mounted on an axle 2 cm. in diameter. A 12-kg. weight hangs on a string wrapped around the axle. What is the acceleration of the weight when it is released and what is the tension in the string?

10. If the weight in Prob. 9 slides down a smooth inclined plane, making an angle of 30° with the horizontal with the string parallel to the plane and rotates the sphere, what will be the acceleration of the weight and the tension in the string?

11. If the inclined plane of Prob. 10 is rough with a coefficient of friction of one-third, what will be the acceleration of the weight and the tension in the string?

12. A mass of 5 kg. is whirled around in a circle in a horizontal plane thirty times a minute at the end of a rod 1 m. long. What is the tension in the rod?

13. A pilot looping the loop nearly falls from his seat at the top of the loop, going 80 mi./hr. What is the radius of the loop?

14. A ball on the end of a string rotates about a vertical post. The string is 4 m. long and is attached to the top of the post. How many revolutions per second must the ball make in order that the string make an angle of 60° with the post?

15. Suppose 10 lead balls are mounted on the ends of wires forming the spokes of a wheel. The balls are each of mass m kg., and their centers are at a distance r m. from the axis of the wheel. If the tension necessary to break each wire is T newtons and the wheel is accelerated by a torque of L newton-m., how long after starting will balls begin to fly off? Assume that the wheel is rotating in a vertical plane, and neglect the mass of the wires and friction.

16. Suppose a mass M_1 on a horizontal table is connected with a mass M_2 hanging over the edge by a string passing over a pulley of radius R and moment of inertia I. Neglecting the friction of the table and pulley bearings and assuming no slipping of the string on the pulley, what is the acceleration of the system?

CHAPTER XIX

ELECTRONS

1. In earlier chapters, we have interpreted various electrical phenomena in terms of electrons and their motions. We have discussed the determination of the electronic charge and have used the fast moving electrons in a beam of cathode rays as an example of an electric current. But we have still to explain why we think that the tiny discrete charges that Millikan measured are associated with particles so small that they can move about in solid conductors, why we think cathode rays are rapidly moving particles of this type, why we think that the carriers of small negative charges of electronic size behave differently from the carriers of small positive charges of the same size, and what role these carriers of positive and negative charge play in the structural scheme of things. We cannot answer all these questions at once, but we can begin to find their answers by making a more careful study of cathode rays than we have made heretofore. We want to study their mass, their charge, their velocity, and their occurrence. Much of this information can be obtained by a quantitative study of the deflections of cathode rays by electric and magnetic fields, the kind of experiments we have already considered in a qualitative way. Consequently, we shall begin by a mathematical study of the effects of electric and magnetic fields on moving charged particles.

Electric Deflection of a Moving Charge

2. Suppose that an ion of mass m and charge e is moving with a velocity v in an evacuated space where its chance of collision with a gas molecule is small. Suppose that for a distance L it passes through an electric field E perpendicular to its original path and then continues a further distance d. We know that the electric field deflects the ion from its original straight line path, and our problem is to find out how great this deflection is and on what it depends. The force on the ion is Ee, in the direction of

the electric field if e is positive, in the opposite direction if e is negative. The path of the particle, therefore, is exactly similar to the path of a projectile in the gravitational field of the earth. We saw that this was a parabola. (We should point out, incidentally, that the effect of gravity on the positive and negative ions we shall be talking about is so small compared with the electric and magnetic effects that it can be entirely neglected.)

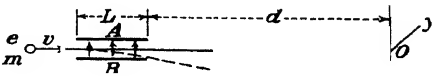


FIG. 177.—Deflection of an electrically charged particle by an electric field. The charge is taken as negative.

Let us choose x and y coordinate axes in a plane perpendicular to the plane of the paper in Fig. 177 so that the x-axis is parallel to the direction of the electric field. Then the distance that the ion will be deflected in passing between the plates A and B will be $X = at^2/2$. But the time t that the ion is between the plates is $t = L/v$ and the acceleration is Ee/m so that the distance that the ion travels vertically while going through the plates is $X = \frac{1}{2}(EeL^2/mv^2)$. Furthermore, in that time, it acquires a

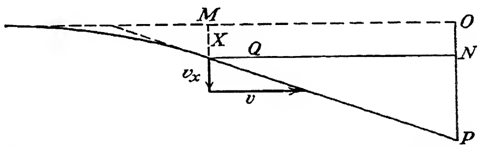


FIG. 178.—Detail of the deflection of a negatively charged particle by an electric field. X represents the deflection while the particle is in the field, and NP the deflection after the particle has passed through the field.

vertical component of velocity $v_x = at$ or EeL/mv while maintaining its original horizontal velocity v.

3. After emerging from between the plates, the ion is not accelerated by any further force and continues in a straight line with both its horizontal and vertical components of velocity unchanged. It will travel the horizontal distance d to the fluorescent screen in a time d/v and in that time will move an additional distance downward $v_x(d/v)$ or $EeLd/mv^2$. Therefore,

if we let x be the total deflection OP of the fluorescent spot, we have in Fig. 178

$$x = ON + NP = X + NP = \frac{EeL^2}{2mv^2} + \frac{EeLd}{mv^2}$$

or

$$= L\left(\frac{L}{2} + d\right)\frac{Ee}{mv^2}. \tag{1}$$

From this equation, we see that the *deflection* of a moving charged particle by an *electric field* is inversely proportional to the *kinetic energy of the particle.* It depends also on the charge on the particle, the strength of the field, and the dimensions of the apparatus, as might be expected. All the quantities appearing in Eq. (1) can be determined directly except e, m, and v. We might assume e to be the electronic charge already determined by Millikan's experiment, but we would still need some other experiment to determine v or m.

Magnetic Deflection of a Moving Electric Charge

4. In Chaps. XIII and XIV, we saw that a moving electric charge was equivalent to an electric current. It must, therefore, be acted upon by a magnetic field and, indeed, we have seen that a cathode-ray beam is deflected by a magnetic field. More specifically, we saw that a beam of charged particles moving with a velocity v and each carrying a charge e was equivalent to a current nev if n was the number of particles per unit length of the beam. Therefore, we can find the deflecting force of a magnetic field on such a beam by applying the law $\mu HI \sin\theta$ for the force on a unit length of current I in a magnetic field H. Substituting nev for I and *assuming that the field is perpendicular to the direction of motion of the ions*, we have $\mu_0 Hnev$ as the force per unit length of the beam where $\mu_0 = 4\pi \times 10^{-7}$ newton/amp.^{2} is the permeability of empty space as is appropriate in this case. But n is the number of ions per unit length of the beam; therefore, the force per ion is simply

$$F = \mu_0 Hev \tag{2}$$

and is perpendicular to both the direction of motion of the ion and the direction of the magnetic field.

5. Apply this result to the case of a beam of ions like that of Par. 2, substituting a magnetic field perpendicular to the plane of the paper for the electric field in the plane of the paper. As shown in Fig. 179, the result is similar. There is, however, a significant difference. In the electric field the force on the charge is independent of the direction and magnitude of the charge's velocity, but in the magnetic field the direction of the force is always perpendicular to the direction of the velocity of charge and proportional to its magnitude. We have just been discussing in the last chapter a case of motion where the only force acting is always perpendicular to the direction of the motion. There we found such a situation produced motion in a circular path. We can conclude at once, therefore, that a charged particle moving in a uniform magnetic field travels in a circular path. Of course, in an arrangement such as is shown in Fig. 179, the particle passes out of the field after traveling through only a small part of the circle. Nevertheless, the curvature of the trajectory between T and Q', the limits of the magnetic field, is given by setting the centripetal force equal to $\mu_0 Hev$. This gives us

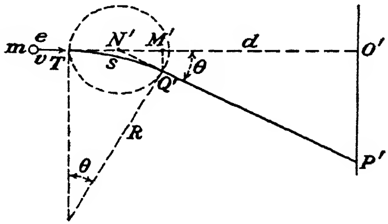


FIG. 179.—The deflection of an electrically charged particle by a magnetic field. The field is directed into the paper if the charge on the particle is taken as negative.

$$\frac{mv^2}{R} = \mu_0 Hev$$

or

$$R = \frac{mv}{\mu_0 He} \tag{3}$$

where R is the radius of curvature of the path of the ion in the field. If s is the distance that the ion travels in the magnetic field and θ is the angle between the trajectory coming out of the field and the original trajectory, then θ is a measure of the deflection caused by the magnetic field and is equal to s/R. This gives us from Eq. (3)

$$\theta = \frac{s\mu_0 He}{mv} \tag{4}$$

which is a perfectly general and exact result. It shows that the *deflection* of a *moving charged particle* by a *magnetic field* is inversely proportional to the *momentum of the particle.* As might be expected, it depends also on the strength of the magnetic field, the charge on the particle, and the dimensions of the apparatus.

6. To get from Eq. (4) to a relation giving the displacement of the fluorescent spot on the screen, we shall make the assumption that the angle θ is very small. Then s can be taken as equal to L, the length of the region in which there is a magnetic field; and $Q'M'/M'N'$, which is strictly equal to the tangent of θ, can be taken as equal to θ itself. We then have from similar triangles that $O'P'/O'N'$ equals $Q'M'/M'N'$ equals θ. $O'P'$ is the displacement of the fluorescent spot which we are trying to get. Setting $O'P' = y$, we have

$$y = (O'N')\theta = O'N'\frac{s\mu_0 He}{mv};$$

but, if θ is very small,

$$s = L \qquad \text{and} \qquad O'N' = d + \frac{L}{2};$$

therefore,

$$y = L\left(\frac{L}{2} + d\right)\frac{\mu_0 He}{mv}. \tag{5}$$

We see that this is similar to the equation for the electric deflection except that the magnetic induction $\mu_0 H$ appears instead of the electric field strength and the momentum of the particle appears instead of twice the kinetic energy. It is to be remembered also that in this case the deflection is perpendicular to the magnetic field, whereas in the electric case it was parallel to the electric field, also that this equation is true only for small deflections.

7. The avowed object of the first part of this chapter is to explain how the mass and velocity of cathode rays are determined. Actually, the preceding discussion of electric and magnetic deflection is fuller than is necessary for this purpose, but there are so many other purposes for which these deflections are useful that it seemed desirable to treat them fully at this point in order to avoid going over the subject again later on.

We can, for example, make use of the preceding equations in explaining the "mass spectrograph" and the "cyclotron." We shall now proceed with the particular application of the method to cathode rays.

The Velocity of Cathode Rays

8. Suppose that the actions of the electric and magnetic fields treated in the foregoing paragraphs are combined. That is, suppose that when the cathode-ray beam enters the region between the two plates that are setting up the electric field it simultaneously enters a region between two poles of a magnet that set up a magnetic field perpendicular to the electric field. The deflecting forces are then Ee and $\mu_0 Hev$ and are parallel to each other. By adjusting the sign and magnitude of one field relative to the other, these forces can be made equal and opposite so that the cathode-ray beam is not deflected at all and the fluorescent spot it makes on the screen returns to the place where it was when neither field was acting. For this to happen, we must have $Ee = \mu_0 Hev$ or

$$v = \frac{E}{\mu_0 H}. \tag{6}$$

This, therefore, gives us a method for measuring the velocity of the cathode rays.

9. We can get some idea of the order of magnitude of these velocities by considering experiments that can be done in the lecture room. It is found that the fluorescent spot of a cathode-ray oscillograph is deflected a centimeter or two by a potential of 120 volts on plates a few millimeters apart, *i.e.*, a field strength of perhaps 25,000 to 30,000 volts/m., acting for about 1 cm. (=0.01 m.) of the path of the beam. The magnetic fields needed to produce similar deflections act over a larger part of the path and are at most equal to about 1,000 amp. turns/m., *i.e.*, a magnetic induction $\mu_0 H$ of 10^{-3}. To get an equivalent deflection with a magnetic field confined to the region of the electric deflecting plates might require as much as ten times as great a magnetic field. Even if these guesses are very far from exact, it is clear that the velocities of the cathode rays must be very large.

10. Exact measurements give values for the velocities of cathode rays ranging from 1.2×10^7 m./sec. for those used in the cathode-ray oscillograph through 7.3×10^7 m./sec. for those obtained with a 15,000-volt induction coil to upper limits of 1.66×10^8 m./sec. (10^5 volts) obtained with the most powerful transformers. This last value is well up toward the velocity of light. So much so that the apparent mass of the cathode rays increases and the theory of relativity has to be introduced to account for the results.

Ratio of Mass to Charge for Cathode Rays

11. Once the velocity of the cathode rays has been determined by balancing the effects of the electric and magnetic fields, the effect of either one separately can be measured. The measured deflection and the previously determined velocity can then be substituted in the appropriate Eq. (1) or (5) in which there remain but two unknowns e and m. The ratio m/e is one that appears not only in Eqs. (1) and (5) but in many other relations of atomic physics. Moreover, it was accurately determined before it was possible to measure either e or m separately. Its value is

$$\frac{e}{m} = 1.76 \times 10^{11} \text{ coulombs/kg.}$$

or (7)

$$\frac{m}{e} = 5.68 \times 10^{-12} \text{ kg./coulomb.}$$

12. We have tacitly assumed in all this discussion that cathode rays are moving negatively charged particles. Such an assumption is justified by the experiments described in Chap. XIV, Par. 4, as well as by many others which we have not had time to consider. We now make the further assumption that each one of these particles carries the smallest possible charge, that of an electron, and, in fact, we call any small particles having this particular negative charge, and ratio of mass to charge, electrons. Since we know the electrónic charge to be 1.60×10^{-19} coulomb (see Chap. XII), we can find the mass of an electron by substituting this value in (7). By doing this, we obtain

$$m = 9.09 \times 10^{-31} \text{ kg.} \tag{8}$$

It is of interest to note that this is about 1/1,848 of the mass of a hydrogen atom. In short, an electron is a very light particle.

13. The ratio of charge to mass has been measured for the cathode rays obtained from a great many different sources. No change in the nature of the gas in the discharge tube, the material of the electrodes, or of the tube itself has ever been shown to alter the e/m of the cathode rays. Apparently, the negatively charged particles that make up these rays and which we call electrons are universal constituents of matter. In Pars. 15 and 17, we shall see that particles of the same sort are released from the surface of solids under the action of light or high temperature. Correspondingly light positively charged particles are never observed under similar conditions although they have recently been found emitted by certain kinds of artificial radioactive substances and have been christened positrons. The positively charged particles that occur in gas discharges are much heavier and more slow moving than electrons (see the next chapter).

The Kinetic Energy of an Electron

14. The velocity of cathode rays can also be estimated directly from the way in which they are produced. In Par. 3, Chap. XIV, it was mentioned that most of the potential drop in a discharge tube was close to the cathode, *i.e.*, the field strength there was very strong. But by definition, the potential drop between two points A and B is the work required to move a unit charge from B to A or, conversely, it is the energy acquired by a unit charge in moving from A to B under the accelerating action of the electric field. In mathematical language, a charge e falling through a potential drop V acquires a kinetic energy given by

$$\frac{mv^2}{2} = Ve \tag{9}$$

where m is the mass of the particle carrying the charge in kilograms, v is its final velocity in meters per second, e is the charge in coulombs and V is the potential drop in volts. Acceleration by an electric potential drop is so universal a method for giving energy to charged particles and the amount of the potential drop is so much easier to measure than the velocity of the par-

ticle that it has become customary to speak of particles having energies of so many "volts." Strictly, this is an abbreviation for "equivalent electron volts" and means that the particle has as much energy as a particle carrying one electronic charge would have acquired in falling through a potential difference of so many volts. On the basis of this definition,

$$\text{One equivalent electron volt} = 1.60 \times 10^{-19} \text{ joule.} \qquad (10)$$

Energies corresponding to higher voltages are proportional. It may also be interesting to give some values of electron velocities, which vary as the square root of the voltage. The following values are given to two places only and are not corrected for the relativity increase in mass at high velocities.

Voltage	Corresponding Electron Velocity, m./sec.
1.0	5.9×10^5
10.0	1.9×10^6
100.0	5.9×10^6
1,000.0	1.9×10^7

The Thermionic Effect

15. We have used the notion of electrons in talking about the conduction of electricity through metals and also in talking about gas discharges and electrolysis. If there really is such a thing as an electron, there should be some connection between those that appear in the cathode-ray beam of a low-pressure gas discharge and those that carry the current in a metallic conductor. We have implicitly assumed that they are identical. It is easily shown that this is so by considering what is known as the thermionic effect or sometimes the Edison effect. Every radio tube depends on this effect for its operation, but the type that demonstrates it most simply is a two-element rectifier tube (Fig. 180). Such a tube consists of a metallic filament mounted opposite a flat plate in an evacuated glass bulb. The filament can be heated electrically. If the plate is made positive with respect to the filament and the circuit completed through a galvanometer, no current is shown by the galvanometer so long as the filament is cold, but when it is heated current begins to flow. The amount of current flowing is found to increase with the temperature of the filament and with the voltage between the filament and the plate.

It depends also on the material of which the filament is made. If the voltage is reversed so that the filament is positive with respect to the plate, no current flows, whatever the temperature of the filament and the amount of the voltage. These results are immediately understandable if we assume that there are electrons in the filament that can escape from it by a process similar to evaporation. The number that evaporate, the electronic vapor pressure, so to speak, increases with the temperature. If the plate is positive with respect to the filament, the negatively charged electrons are drawn across to it. If it is negative, they are driven back toward the filament and no current flows.

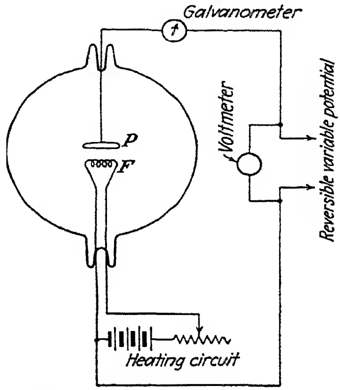


FIG. 180.—A hot-filament vacuum rectifier for showing the thermionic effect. If the filament F is hot and the plate P positive, the galvanometer shows a deflection.

16. It is natural to assume that these electrons which emerge from a metal when it is heated are the electrons that carry the current inside the metal. To show that they are identical with those knocked out of the gas molecules in a low-pressure discharge, we have only to measure their mass and their charge by the methods that have already been discussed. The results of such experiments prove this identity beyond question and are now taken so much for granted that we have used the electrons from hot filaments and from gas discharges quite indiscriminately in illustrating the properties of electrons.

Photoelectric Effect

17. One of the most interesting and useful ways of getting electrons out of metal surfaces is by the action of light or x-rays on them. This phenomenon will be discussed later (Chap. XXVII, Pars. 9–11) after we have studied electromagnetic radiation. The electrons obtained in this way are, like thermionic electrons, usually of low velocity, but they have the same ratio of mass to charge as the electrons in a cathode-ray beam.

SUMMARY

Valuable information about cathode rays is obtained from a quantitative study of their deflection by electric and magnetic fields. An analysis of the deflection of a rapidly moving charged particle by an electric field perpendicular to its direction of motion shows that the deflection is proportional to the charge of the particle and to the electric field strength and inversely proportional to the kinetic energy of the particle. Similar analysis of the deflection by a magnetic field perpendicular to the direction of motion shows that the deflection is proportional to the magnetic field strength and the charge of the particle and inversely proportional to the momentum of the particle. The particle moves in the arc of a circle of radius $R = mv/Be$ where B is the magnetic induction equal to $\mu_0 H$. The circle is in a plane perpendicular to the direction of the magnetic field. By combining the effects of electric and magnetic fields, it is possible to determine the velocity and the e/m of moving charged particles.

The ratio e/m for cathode rays from every kind of source is 1.76×10^{11} coulombs/kg. For well-established reasons, cathode rays are believed to be small particles each carrying one negative electronic charge (16×10^{-20} coulomb) and are identified with what we have been calling electrons. The mass of an electron is, therefore, 9.09×10^{-31} kg., about 1/1,848 the mass of a hydrogen atom.

For cathode rays, the velocities prove to be very high, of the order of 10^7 m./sec. if the potential drop across the discharge tube is of the order of several thousand volts. In fact, these velocities can be calculated directly in terms of the kinetic energy acquired by an electron falling through a potential difference V since $Ve = mv^2/2$.

Electrons are emitted by hot bodies (thermoelectric effect) and under the influence of x-rays or ultraviolet light (photoelectric effect).

ILLUSTRATIVE PROBLEMS

1. A proton, *i.e.*, a hydrogen atom carrying a single positive electronic charge, traveling 10^6 m./sec. passes for a distance of 20 cm. through an electric field of 10^4 volts/m. at right angles to its path. If the proton strikes a screen placed at right angles to its original path 40 cm. from the exit edge

of the electric field, what will be its deflection from the point where it would have struck the screen with no field present?

The deflection of a moving charged particle by an electric field at right angles to its velocity is [Eq. (1), page 390]

$$x = L\left(\frac{L}{2} + d\right)\frac{Ee}{mv^2}$$

where (Fig. 177, page 389),

L = 0.20 m. = the length of path through the field.
d = 0.40 m. = the distance from the field to the screen.
E = 10^4 volts/m. = the electric field strength.
e = 1.60×10^{-19} coulomb = the charge on a proton.
m = 1.66×10^{-27} kg. = the mass of a proton.
v = 10^6 m./sec. = the velocity of the proton.

Substituting these values in Eq. (1) gives

$$x = 0.20 \text{ m.}\left(\frac{0.20 \text{ m.}}{2} + 0.40 \text{ m.}\right)\frac{10^4 \text{ volts/m.} \times 1.60 \times 10^{-19} \text{ coulomb}}{1.66 \times 10^{-27} \text{ kg.} \times (10^6 \text{ m./sec.})^2}$$
$$= \frac{0.2 \text{ m.} \times 0.5 \text{ m.} \times 1.60 \times 10^{-15} \text{ volt coulomb/m.}}{1.66 \times 10^{-15} \text{ kg.-m.}^2/\text{sec.}^2}$$
$$= \frac{0.20 \times 0.5 \times 1.60}{1.66}\,\frac{\text{joules}}{\text{kg.-m./sec.}^2} = 0.0964 \text{ m.} = 9.64 \text{ cm.}$$

2. A singly charged sodium ion traveling 10^5 m./sec. passes through a circular magnetic field 2 cm. in diameter of 10^4 amp. turns/m. at right angles to its path. If the ion strikes a screen placed at right angles to its initial path 30 cm. from the edge of the magnetic field, what will be its deflection from the point where it would have struck the screen with no magnetic field present?

The deflection of a charged ion in a magnetic field is given by Eq. (5), page 392, as

$$y = L\left(\frac{L}{2} + d\right)\mu_0\frac{He}{mv}$$

where from Fig. 179, page 391,

L = 0.02 m. = the diameter of the magnetic field.
d = 0.30 m. = the distance from the magnetic field to the screen.
μ_0 = $4\pi \times 10^{-7}$ newton/amp.^{2} = the permeability of empty space.
H = 10^4 amp. turns/m. = the strength of the magnetic field.
e = 1.60×10^{-19} coulomb = the charge on the sodium ion.
m = $\dfrac{0.023 \text{ kg. atoms/mole}}{6.03 \times 10^{23} \text{ atoms/mole}} = 3.81 \times 10^{-26}$ kg. = the mass of a sodium atom, which neglecting the mass of the electron removed, is the mass of the sodium ion.
v = 10^5 m./sec. = the velocity of the ion.

Substituting these values in Eq. (5) gives

$$y = 0.02\text{ m.}\left(0.3\text{ m.} + \frac{0.02\text{ m.}}{2}\right)\frac{4\pi \times 10^{-7}\text{ newton/amp.}^2 \times 10^4\text{ amp. turn/m.}}{3.81 \times 10^{-26}\text{ kg.}} \times \frac{1.60 \times 10^{-19}\text{ coulomb}}{10^5\frac{\text{m.}}{\text{sec.}}}$$

$$= \frac{0.02 \times 0.31 \times 4\pi \times 1.60 \times 10^{-22}\text{ m.}^2\text{ newton-coulomb/amp. m.}}{3.81 \times 10^{-21}\text{ kg.-m./sec.}}$$

$$= 3.27 \times 10^{-3}\frac{\text{kg.-m.}}{\text{sec.}^2}\,\frac{\text{coulomb-sec.}^2}{\text{coulomb-kg.}} = 3.27 \times 10^{-3}\text{ m.} = 0.327\text{ cm.}$$

PROBLEMS

1. An electron traveling 10^6 m./sec. passes for a distance of 10 cm. through an electric field of 10^3 volts/m. at right angles to its path. If the electron strikes a screen 30 cm. from the edge of the electric field, what will be its deflection from the point where it would strike the screen with no field present?

2. Find the equation of the path of an ion in an electric field perpendicular to its path. Write x and y in terms of the time t, and then eliminate the time.

3. A proton moves with a velocity of 10^7 m./sec. at right angles to a magnetic field of 10^5 amp. turns/m. What force acts on it, and what acceleration does it receive?

4. A singly charged fluorine ion is deflected by a magnetic field of 10^4 amp. turns/m. at right angles to its path. The original velocity of the ion is 10^4 m./sec. and it travels 2 cm. through the magnetic field. Find the angle by which it is deviated.

5. Derive the equation for y, the deflection of a moving charged particle by a magnetic field, without assuming θ to be a small angle. [Equation (5) page 392, is the corresponding equation based on this assumption.]

6. An electron traveling 10^6 m./sec. passes through a circular magnetic field 2 cm. in diameter of 100 amp. turns/m. at right angles to its path. If the electron strikes 30 cm. from the edge of the magnetic field, what will be its deflection from the point where it would have struck the screen with no magnetic field present?

7. How much error is made in Prob. 6 by assuming the angle of deviation to be small (*cf.* Prob. 5.)

8. An electron beam passes through a magnetic field of 10^4 amp. turns/m. perpendicular to an electric field of $4\pi \times 10^4$ volts/m. No deflection of the electrons is produced. What is the velocity of the electrons?

9. How many volts difference of potential are required to give an electron a velocity of 10^6 m./sec.?

10. An electron accelerated by a difference of potential of 10^4 volts passes through crossed electric and magnetic fields without being deflected. What is the ratio of the magnetic to the electric field?

CHAPTER XX

POSITIVE RAYS AND ISOTOPES

1. The probability that all matter consisted of small particles called atoms was established in the first part of this book by a consideration of chemical evidence and of the successes of the kinetic theory. Then, in the study of electricity, it was suggested that the current might be carried by atoms or parts of atoms each carrying a definite amount of charge. The laws of electrolysis and Millikan's experiment were used to establish this notion of the atomicity of electric charge on a firm basis. This idea was used frequently in the treatment of the phenomena of current electricity and magnetism when current was interpreted in terms of the motion of millions of tiny particles called electrons, each having a negative charge of 16×10^{-20} coulomb. In the last chapter, we returned to a more direct study of these electrons. A renewed study of gas discharges has shown that electrons are identical with cathode rays and has given a measurement of the ratio of their mass to charge. Their presence in metals has been proved directly by thermionic experiments. Since the value of the charge they carry was already known, their mass could immediately be found. Furthermore, the mass and charge of an electron are entirely independent of the material from which it comes although there is no material known from which it is impossible to obtain electrons.

2. Matter is known to be varied in nature, heavy, and electrically neutral under most circumstances. Apparently, electrons are universal constituents of matter but they are always the same, very light, and negatively charged. Consequently, there must be other constituents of matter which give variety to the different elements, give them most of their mass, and carry enough positive charge to neutralize the negative charges carried by the electrons. Let us fix our attention on this last point for the moment. If the molecules of the gas in a discharge tube are neutral before the current is turned on and then enough

electrons are removed from them to form the beam of cathode rays that we have observed, many of the gas molecules must be left with positive charges. Such molecules from which some electrons have been removed are called positive ions like the similar charged molecules in electrolysis. They are called ions (from the Greek verb meaning go) because they move under the action of the electric field in the tube. They go toward the cathode, the negative electrode. In doing so, they are continually accelerated by the electric field and acquire considerable momentum. If the cathode has a number of holes in it, some of the ions will go through these holes into the space beyond and cause luminosity by collisions with the gas molecules there. This effect is actually observed.

(It may seem strange that the cathode in a discharge tube plays so much more important a part than the anode in the study of both positively and negatively charged ions. We talk of the cathode rays as being shot away from the region in front of the cathode and of the positive ions as being attracted toward the cathode as if the location and nature of the anode made very little difference. This is found to be true, a fact that arises from the difference between the natures of the positive ions and the electrons themselves. This difference produces a great asymmetry in the distribution of the potential drop between the electrodes, an asymmetry so marked that almost the entire potential drop occurs within a few centimeters from the surface of the cathode. Consequently, the forces acting on both positive and negative ions are concentrated in this region and are relatively small elsewhere in the discharge tube. The detailed explanation of this effect is not entirely clear and is much too complicated to treat here.)

3. It was natural to apply to the study of positive rays the methods that had been so successful with cathode rays, but the problem proved to be more difficult. Magnetic fields that curled cathode rays into spirals had but slight effect on positive rays, and for a long time no appreciable deflection was obtained with electric fields. This difficulty proved to be caused by the comparatively great mass of the positive rays. Finally, Sir J. J. Thomson solved the problem in a series of experiments so important that they will be described in detail.

The Parabola Method

4. The apparatus used by J. J. Thomson is shown in the drawing below. In the large spherical discharge tube A there was a cathode B through which a very small hole had been drilled. Some of the positive ions attracted to the cathode shot through this hole and on between the plates P and P' which acted as condenser plates to produce an electric field. These plates also served as the pole pieces of a large electromagnet which produced a magnetic field parallel to the electric field. The beam of ions

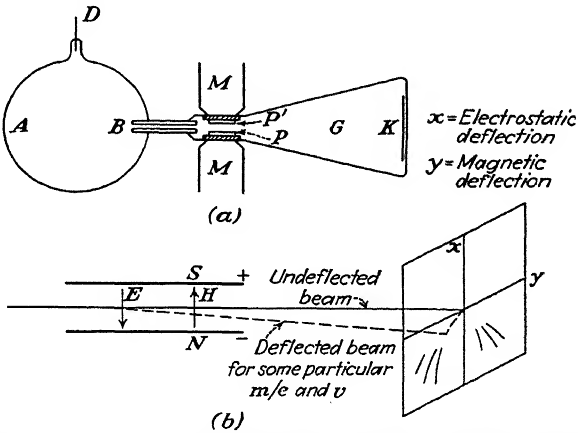


FIG. 181.—J. J. Thomson's positive-ray apparatus. Positive rays from the discharge tube A pass between the electric and magnetic deflecting plates PP' and strike the photographic plate K.

then passed into the space G and struck the photographic plate K. The hole in the cathode was made small to define a sharp beam of positive ions and to make the flow of gas from the discharge tube into the highly evacuated space G as small as possible. If neither the electric nor magnetic fields were on, the ions all struck the photographic plate in the same place, making a large spot; if just the electric field was on, the ions were bent up or down; if just the magnetic field, they were bent in or out.

5. To investigate these deflections more completely, we shall apply the equations developed in Pars. 2 to 6 of the previous chapter. According to that treatment, if L is the length of the

plates P and P' which apply to both electric and magnetic fields and d is the distance from these plates to the photographic plate K, the electric and magnetic deflections are given by

$$\text{Electric deflection} \quad x = L\left(\frac{L}{2} + d\right)\frac{Ee}{mv^2}.$$

$$\text{Magnetic deflection} \quad y = L\left(\frac{L}{2} + d\right)\frac{\mu_0 He}{mv}$$

where e, m, and v are the charge, mass, and velocity of the particle, E is the strength of the electric field, H that of the magnetic field, and μ_0 is the permeability of empty space. We can write the preceding equations in the form

$$x = k\frac{Ee}{mv^2} \tag{1}$$

and

$$y = k\frac{\mu_0 He}{mv} \tag{2}$$

where $k = L\left(\frac{L}{2} + d\right)$ and is constant for a given apparatus.

6. Consider first the electric deflection, shown as vertical in Fig. 181. If we confine ourselves to ions of the same charge, it is evident that their deflection is inversely proportional to their kinetic energy. But this kinetic energy is acquired in the discharge tube A. Ions there are accelerated by the electric field between D and B and gain kinetic energy. In general, they will lose some of this energy in collisions; but even if the pressure is so low that many of the ions make no collisions, the greatest possible energy they can acquire is that corresponding to the whole potential drop between D and B, *i.e.*, Ve where V is the potential drop across the discharge tube and e is the charge on the ion. To this maximum energy, there corresponds a minimum electrostatic deflection which is shown clearly on the photographs (Fig. 182) by the sharp cutoff of most of the parabolas along a vertical line to the right of the origin.

7. There is no corresponding maximum momentum to set a definite limit to the magnetic deflection. For a given kinetic energy, the velocity will be large for ions of small mass, smaller for the heavy ions. For a given mass and charge, the velocity

will vary downward from a maximum for ions of kinetic energy = Ve. This variable velocity is of less interest to us than the m/e ratio of the ions.

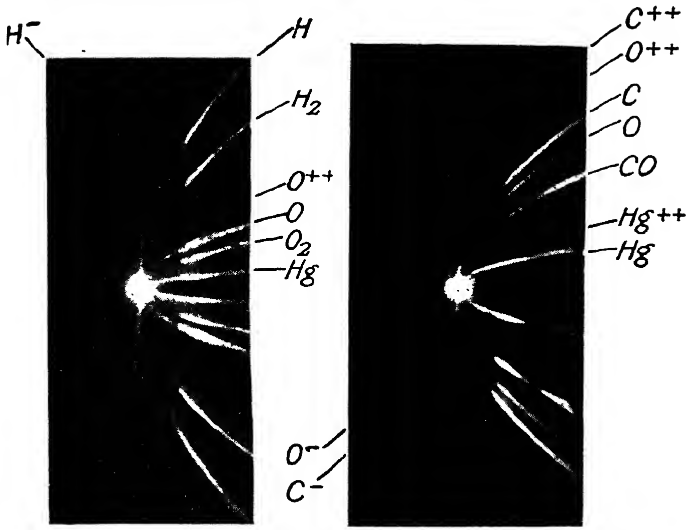


FIG. 182.—Photographs of the typical positive-ray parabolas. The electric deflections are horizontal and the magnetic deflections vertical. The bright spot is the origin. The parabolas show a minimum value of the electric deflection corresponding to the maximum energy Ve of the rays.

8. To bring out the effect of the m/e of the ions on the relative magnetic and electric deflection, we eliminate v from Eqs. (1) and (2) by squaring (2) and dividing which gives us

$$y^2 = k\frac{\mu_0^2 H^2}{E}\frac{e}{m}x \tag{3}$$

or

$$y = \pm\sqrt{k\frac{\mu_0^2 H^2}{E}\frac{e}{m}x}$$

This is the equation of a parabola with its vertex at the origin and its axis along the x-axis. Thus, all ions of the same value of m/e will strike the photographic plate somewhere along a particular parabola, the exact point depending on their velocity.

The greater the value of m/e, the closer the parabola will be to the x-axis. If there were a continuous variation in the mass of the ions, there would be a continuous family of parabolas and a vague blur on the photograph. It is clear from the picture given that this is not the fact. (In making these photographs, it is customary to reverse the magnetic field in the middle of an exposure so that the parabolas appear in more than one quadrant.) Since only a finite number of parabolas appear, evidently there are only a finite number of m/e values represented by the ions in the discharge. These values can be determined by a knowledge of the strengths of the electric and magnetic fields and the dimensions of the apparatus.

9. We have seen that the atomic weight of an element is proportional to the average mass of an atom of a particular chemical species and is numerically equal to the weight in grams of one mole of that element. It is convenient in the kind of work we are now discussing to speak of the masses of individual atoms in terms of the average weight of the atoms of some particular element. The element so chosen is oxygen, and the average mass of the atoms of oxygen is said to be 16.0000. Since we know that the mass of an atom of oxygen is, on the average, $16/6.03 \times 10^{23}$ g., it is clear that the unit of mass in this system is

$$\frac{1}{6.03 \times 10^{23}} \text{ g.}$$

In this system, the average mass of a hydrogen atom is 1.008 and may be taken as 1.0 for most of our purposes. The average mass of an atom of any other element in this system of units is equal to the atomic weight.

10. These figures are obtained from chemical evidence. They do not necessarily give the mass of any single atom. The individual atoms might have any one of several masses or even a continuous range of masses distributed about a most probable value as far as chemical evidence is concerned. To be sure, the simplest assumption is that they all have the same mass, but it was the experiments on positive rays that first threw light on the question. At first sight, it appeared that the simplest assumption was the right one. The plates showed a small number of fairly sharp parabolas each one of which corresponded to a different atom or molecule with one or two electronic charges. (Through-

out this discussion, the electronic charge 16.0×10^{-20} coulomb is taken as the unit of charge. In the last paragraph, we chose to base our unit of mass for this kind of discussion on the atomic weights. In these mass and electronic charge units, the average ratio of mass to charge for atomic ions carrying single electronic charges will be the atomic weight of the element. Thus, the average m/e for hydrogen atomic ions is about one, the average m/e for Hg^+ ions is 200.6, for Hg^{++} ions $200.6/2 = 100.3$, and so on.) But in a series of experiments in which Aston studied the rare gas neon, there always appeared two parabolas, one about ten times as strong as the other and neither of them coming at just the right place for ions of $m/e = 20.2$. One of those parabolas, the strong one, came at $m/e = 20$, the other at $m/e = 22$. As a result of this and other experiments, the conclusion became inescapable that neon was a mixture of at least two kinds of atoms chemically identical but differing in mass. Such atoms are called isotopes.

11. To anyone familiar with the years of painstaking work that chemists had spent in making exact determinations of the atomic weights of the elements, it will not be surprising that the discovery of isotopes was revolutionary in its importance. The atomic weight was immediately dethroned from its position of high significance, though retaining its indispensability for calculations of chemical reactions. As we shall see later, there had already been indications from the studies of x-rays and radioactivity that the atomic number* was of more importance in determining the nature of an element than the atomic weight, but evidence was now at hand that atoms could differ in weight and yet be chemically identical. Nor was this the only important implication of Aston's early experiments. Not only had he found that neon was made up of two kinds of atoms, but he had found that the weights of these two kinds of atoms were 20.0 and 22.0, exactly whole numbers, although the atomic weight of neon was 20.2, not a whole number. This result carries us back more than a hundred years to a suggestion made by an English physician named Prout. At that time, only a few atomic weights were known and they none too accurately. In studying them, Prout was struck by the fact that they all were whole numbers or very

* Given in the tables of elements at the back of the book. For the definition, see Chap. XXI, Par. 10.

nearly so. He guessed that all the true atomic weights were exactly whole numbers, and to explain this fact he set up the hypothesis that all the chemical elements were made up of hydrogen which had the atomic weight one. In the course of time, methods of chemical analysis improved and atomic weight determinations became so accurate that it was clear that the atomic weights were not whole numbers. Chlorine for example has an atomic weight of 35.5 and, as we have seen, neon has an atomic weight of 20.2. Consequently, Prout's hypothesis was discarded as disproved. The discovery of isotopes with integral atomic weights revived it.

12. The suggestion that all matter is made of one primordial substance is very satisfying philosophically. So much so that it had been made by philosophers long before Prout's time without any substantial experimental evidence to support it. Naturally, the experimental test of such a suggestion was a matter of great interest. Nor is the measurement of the masses of atoms the only test that suggests itself. If all elements are made of hydrogen, the age-old dream of the alchemists ceases to be a mere idle fancy; the notion of changing one element into another becomes very reasonable. Such changes have been effected in recent years for many elements although only on a small scale. To discuss this subject intelligently, we need some knowledge of radioactivity and of electromagnetic radiation, topics that we have not yet studied; but we are already equipped with sufficient information to consider fully the results on isotopes obtained by Aston and others using the methods of positive-ray analysis.

Atomic-mass Analysis

13. The general principles on which all the various methods of atomic-mass analysis depend have been illustrated by the description of the parabola method. The first problem is a source of positive ions. We have as yet no satisfactory method for getting exact measurements of the masses of uncharged atoms. It is necessary that they should be charged so that they can be subjected to the action of electric and magnetic fields. The only force that acts directly on the mass is the gravitational force. This is not under our control and is much smaller than the electric

and magnetic forces we can set up. Our first problem, then, is to get the atoms that we want to study into a state of ionization. There are two general methods for doing this. The first, used by Aston in most of his experiments and also by other investigators, sets up a high-voltage discharge in a low pressure of a gas containing the element to be studied. The second method depends on the fact that certain substances emit positive ions when they are heated; this method is of less general application but is very convenient when it can be used. Once ions have been formed, it is necessary to give them high velocities in a sharply defined direction since the action of the magnetic field depends on the motion of the charges, and since the separation of different kinds of ions depends on small changes in the direction of motion. If the ions are formed in a discharge tube, the high velocities are produced automatically by the large electric fields present; if the ions come from a heated solid source, they must be accelerated by an electric field.

14. We can call the whole arrangement for the formation and acceleration of the ions the source and the rest of the apparatus the analyzer. In general, ions from the source pass into the analyzer through a system of slits which defines a narrow beam. This beam consists of ions of various masses and charges and various velocities. These ions are then passed through some arrangement of electric and magnetic fields which alters the direction of motion of some of them more than others. As we have seen, the action of the electric field depends on the charge and kinetic energy of the ions, whereas that of the magnetic field depends on the charge and momentum of the ions. If the ions all had the same velocity, either a magnetic field or an electric field would be sufficient to separate the ions of different m/e. The detection of the ions after passage through the analyzer can be effected in either of two ways. They can be allowed to strike a photographic plate as in the parabola method, or they can be detected electrically. In the latter procedure, it is customary to measure the current carried by the ions to one particular spot and to vary the strengths of the electric or magnetic deflecting fields so that the ions of different m/e fall successively on that spot. In the photographic method, all the ions are detected at the same time. These principles will now be illustrated by the description of some of the more important apparatuses that have been used.

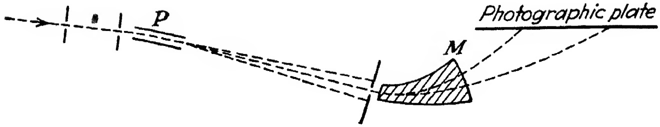


FIG. 183.—Schematic diagram of Aston's mass spectrograph.

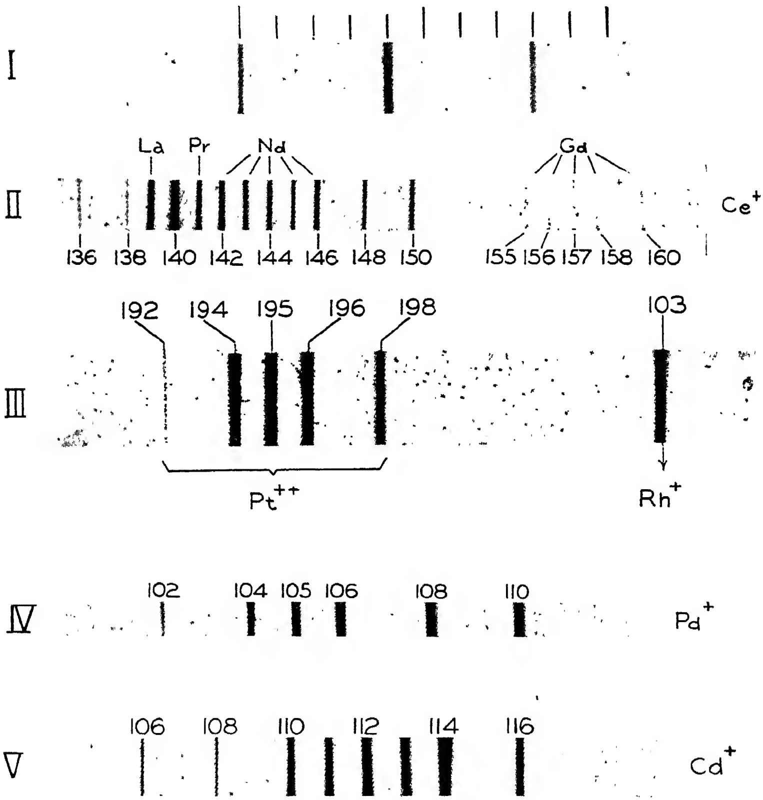


FIG. 184.—Typical mass spectra. These were taken by Dempster but are similar in general to Aston's. The lines in I show the effect of changing the energy of the ions by ±2 per cent. The photograph (IV) of the palladium isotopes should be compared with the electrical record of the same group in Fig. 188.

Aston's Mass Spectrograph

15. Until the last few years, Aston has had almost a monopoly on the determination of the masses and abundance of isotopes. His work has been done with arrangements of the type shown schematically in Fig. 183. Ions of varying energy and m/e enter from the ion source at the left (not shown). The slit system defines a narrow beam which passes through the condenser plates P. An electric field between these plates deflects the ions downward, spreading them out into a fan-shaped beam with the high-energy ions suffering the least deflection. The middle section of this beam then passes into the magnetic field M perpendicular to the paper and in such a direction as to deflect the ions upward again. This latter deflection will depend on the momentum of the ions. Though it is not immediately obvious, it can be shown that the combination of fields in this particular geometrical setup brings all the ions of a particular m/e to the same point on the photographic plate. For this reason, it is said to have a focusing effect; it brings to one spot all the ions of varying velocity which in the parabola method are spread out along a segment of a parabola. This not only gives high intensity, but the record left on the plate by the ions of one kind is a sharp band of the sort shown in Fig. 184. A whole series of such bands usually appears on a single plate just as a series of parabolas appeared on each plate in the parabola method. Because the general appearance of one of these plates is very much like the appearance of a plate on which a spectrum has been photographed, Aston called it a mass spectrum and his apparatus a mass spectrograph. This has proved a convenient term and is now applied to any instrument used for measuring masses by electric and magnetic deflections even when there is no resemblance between the results obtained and an optical spectrum.

Dempster's Method and Its Modifications

16. The apparatus that Dempster set up in Chicago in 1919 is simpler in principle than Aston's and has been used in various modifications by a number of other investigators. However, the modification used by Bainbridge is the most satisfactory one for us to discuss. It is shown schematically in Fig. 185. For the particular form shown, the source of ions is a discharge tube again,

as in Aston's work. The ions emerge with varying velocities but after passage through the slit system S_1S_2 go into a so-called velocity filter. This is a combination of an electric field E in the plane of the paper and a magnetic field perpendicular to it. It will be recalled that for one particular velocity the deflecting force of the electric field will be just equal and opposite to that of the magnetic field. Ions of this velocity will therefore pass through this region in a straight line and emerge through another slit, S_3, into the region where only the magnetic field is present. Ions of any other velocity will be deflected to one side and not get through the lower slit. The ions of suitable velocity to get through the velocity filter need not necessarily be of the same m/e. Those of different m/e will be separated subsequently by the magnetic field, traveling in semicircular paths to different places on the photographic plate. From the discussion in the last chapter, the radii of these paths are evidently given by the equation $R = mv/\mu_0 He$. A typical mass spectrum obtained with this apparatus is shown in Fig. 186.

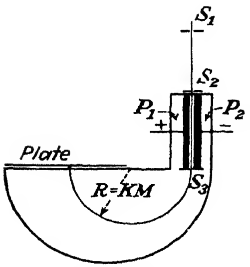


FIG. 185.—Schematic diagram of one of Bainbridge's early mass spectrographs.

FIG. 186.—Mass spectrum of germanium taken by Bainbridge.

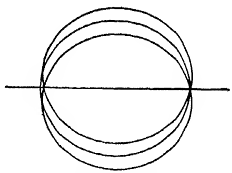


FIG. 187.—Approximate focusing effect of semicircular paths. Ions starting from the common point of the circles on the right come almost together again at the left.

17. Another modification of Dempster's apparatus does not have the velocity filter but imposes the condition of identical kinetic energy by having the ions all fall through the same potential drop between the ion source and the analyzer and making this drop very much larger than any that occurs in the ion source itself. This procedure is particularly satisfactory when the ion

source is an incandescent solid since the ions then have negligible initial energies. The advantage of bending the ions in semicircular paths in the magnetic field is illustrated in Fig. 187, which shows how a slightly divergent beam of ions is brought to an approximate focus. This is caused by the geometrical property that several circles of equal radius but different centers will, if they intersect at one point, come very near to another point of

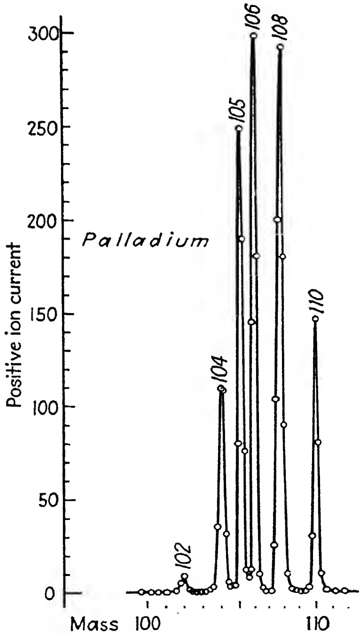


FIG. 188.—The relative abundance of the isotopes of palladium as observed electrically with one of Bleakney's mass spectrographs.

intersection again 180° away. This focusing action has the same advantage that the similar effect has in Aston's spectrograph though its origin is much clearer.

18. Such plates as are reproduced in Figs. 184 and 186 are very satisfactory for the detection of isotopes and the exact determination of their mass, but they are not very satisfactory for the determination of relative quantities. Even by the best methods of measurement of the darkening of the photographic plate, it is

difficult to estimate the ratio of the number of ions making one line to the number making another. For such measurements and for several other purposes for which a mass spectrograph can be used, the electrical method of detection is preferable. In this method, the photographic plate is replaced by a slit behind which is mounted an insulated electrode. This electrode is connected to a very sensitive current-measuring device, usually an "electrometer" tube and amplifier. This measures the current getting through the slit for any given values of the electric and magnetic fields. When no ions are traveling in such paths that they reach the slit, there is no current; but if the electric or magnetic field is gradually altered, the deflections of the ions will alter until those of a particular m/e just get through the slit, carrying charge to the electrode. As the field is still further altered, these ions are deflected still less (or more) and the current decreases again until another type of ion begins to come through the slit, raising the current to another peak. Thus, the presence of ions of several different values of m/e is recorded by the presence of several sharp "peaks" in the curve of ion current against field strength. Figure 188 shows a curve of this type taken on Bleakney's mass spectrograph. In this example, the magnetic field was held constant, and the voltage used to give the ions their initial kinetic energy was varied.

Results of Mass-spectrograph Experiments

19. The results that Aston, Dempster, and others have obtained with various mass spectrographs confirm the implications of the earlier experiments. The masses of all known atoms are approximately whole numbers on the scale that takes the average mass of the atoms of oxygen as exactly 16.0000. This makes it possible to assign a number to any atomic species which we call the *mass number* and which is the whole number nearest to the mass of that particular kind of atom on this scale. Thus, we would say that neon consisted of three species of atoms or had three isotopes of mass numbers 20, 21, and 22. In Table 17, some results of this type of investigation are collected and brought up to date as of July, 1937. The mass numbers are given in the fourth column of the table. It will be seen that oxygen itself has three isotopes of mass numbers 16, 17, and 18. Since the oxygen sixteen isotope is very much more common than the two others,

the average mass of the oxygen atoms is very nearly that of the oxygen sixteen isotope. Consequently, it makes no difference whether we take the average mass of the oxygen atoms in our definition of mass number as we did above or the mass of the sixteen isotope. The whole number nearest to the exact mass will be the same whichever basis is used for the exact mass. But the exact mass itself will differ slightly, and it is necessary in expressing it to specify which is the standard. For reasons that we need not discuss, it has been decided to use the mass of the oxygen sixteen isotope taken equal to 16.0000 as the basis for the scale of masses of the various atomic species. It is on this basis that the figures in column 7 of the table are expressed.

20. It will be seen from the results given in column 7 that the exact mass of an atomic species in no case differs from the mass number by as much as a tenth of a unit. Nevertheless, there is a real difference exceeding the experimental error in nearly every case. The interpretation of this difference is of great interest in its bearing on some of the most fundamental problems of modern physics. The amount of the difference for each isotope is expressed by the "packing fraction" given by the figures in column 6. These numbers are obtained by subtracting the mass number (column 4) from the exact atomic mass (column 7), dividing by the mass number, and multiplying by 10,000. For example, carbon twelve has an exact mass of 12.0036; therefore, its packing fraction is

$$\frac{12.0036 - 12.0}{12.0} \times 10{,}000$$

which is equal to +3.

Equivalence of Mass and Energy

21. What is the bearing of these results on Prout's hypothesis? If all the elements are built up of hydrogen, why are not all the figures in column 7 integral multiples of 1.0081, the mass of hydrogen one? For instance, the mass of helium four is 4.0039, but four times the mass of hydrogen one is 4.0324. If the helium atom is made of four hydrogen atoms,* what has become of

* We shall see in Chaps. XXI and XXVII that Prout's hypothesis is somewhat too simple. The nucleus of a helium atom consists of two protons and two neutrons. Around this nucleus, two electrons circulate at a comparatively large distance.

$4.0324 - 4.0039 = 0.0285$ unit of mass? The answer is that four hydrogen atomic nuclei jammed together in a highly stable helium nucleus do not have so much mass as four hydrogen nuclei in four separate hydrogen atoms. For this reason, the effect is called the "packing effect." To understand this effect, we shall have to make a digression into an apparently unrelated field, the theory of relativity, though we shall make no attempt to present that subject as a whole. In the theory of relativity, it is found that the inertia of a body increases with its velocity. Since mass is defined in terms of inertia, this means that the mass increases with the velocity. Actually, it is found that the mass at a speed v is given by

$$m = \frac{m_0}{\left(1 - \frac{v^2}{c^2}\right)^{\frac{1}{2}}} \tag{4}$$

where m_0 is the mass of the body at rest and c is the velocity of light (3.0×10^8 m./sec.). Expanding this expression in a series, we get

$$m = m_0 + \frac{1}{2}\frac{m_0 v^2}{c^2} + \tag{5}$$

if v is small compared with c. In other words, the motion of a body imparts to it an additional inertia equal to its kinetic energy divided by a constant c^2 so that a change in its mass is equivalent to a change in its kinetic energy or vice versa. By an extension of this idea, energy in any form may be considered equivalent to mass and the laws of conservation of mass and conservation of energy become synonymous. Numerically,

$$1 \text{ kg.} = c^2 \text{ joules} = 9 \times 10^{16} \text{ joules} = 2.5 \times 10^{10} \text{ kw.-hr.},$$

or in terms of the units we have been using recently, one mass unit on the atomic weight scale is equivalent to the energy that an electron would acquire in falling through a potential difference of 931×10^6 volts.* This equivalence of mass and energy may seem a very artificial idea and be viewed with skepticism by the beginner. Nevertheless, it is one of the most important results of the theory of relativity and one which has been repeatedly verified in the last few years.

* *Cf.* Chap. XXVII, Par. 23.

22. We are now in a position to discuss the packing effect. A helium nucleus has less than four times the mass of the hydrogen nucleus. Therefore, to get four hydrogen nuclei from one helium nucleus, mass or its equivalent, energy, must be added. In other words, work must be done on the helium nucleus to break it to pieces. Conversely, if a helium nucleus is formed from hydrogen, energy is released. The processes are analogous to chemical reactions in which heat is absorbed or given off. But the amounts of energy involved are enormously greater. For example, the formation of 1 gram molecule of helium from hydrogen would release 6.13×10^8 kg.-cal. If any way could be found to make such energy available, we would have sources of power of an entirely different order of magnitude from anything now available on the earth. It has been suggested that much of the energy in the stars may come from sources of this kind.

23. The helium nucleus has been used as an example of the packing effect. But it is an effect found in all other stable atoms. In fact, if energy did not have to be supplied to break the nuclei apart, they would break up spontaneously and all disintegrate into hydrogen. The importance of accurate measurements of the packing effect is obvious.

Heavy Hydrogen

24. Any experiment that differentiated between the action of individual atoms rather than depending on the average effect of large numbers might be used for the study of isotopes. For example, the photographs of the tracks of individual atoms obtained in a Wilson cloud chamber (Chaps. XXI, XXVII) have been used in certain instances to measure the masses of atoms. The method of measuring molecular velocities described in Chap. II might conceivably be refined sufficiently to measure masses on the kinetic theory assumption that they were inversely proportional to the square of the velocity. Still another method that has actually been very fruitful is the study of the nature of the light given off by different gases in an electric discharge. This method which, more specifically, depends on the study of the band spectrum of a gas, was responsible for the most important discovery of an isotope that has been made, namely, the discovery of hydrogen of mass number two. The advantage of the band-spectrum method is its extreme sensitivity. By studying

the spectrum of hydrogen, Urey, Brickwedde, and Murphy were able to show that there was a small trace of a hydrogen isotope of mass two mixed with the familiar atoms of mass one. This discovery was confirmed by Bleakney using one of the mass spectrographs in Palmer Laboratory at Princeton. These experiments on hydrogen are complicated by the fact that not only atomic ions but also diatomic and triatomic ions appear in the mass spectrum so that any mass spectrum of hydrogen has lines at m/e values of one, two, and three. The way in which this difficulty can be surmounted is somewhat involved and will not be discussed here. It is sufficiently clear cut to make the results perfectly definite for hydrogen two. The heavy isotope of hydrogen is so important that it has been given a name of its own, deuterium.

Separation of Isotopes

25. The attempt to separate the isotopes of an element began almost as soon as they had been discovered. Since the chemical properties of two isotopes of the same element are by definition identical, they cannot be separated by any ordinary chemical means. Their only difference is one of mass, and therefore they must be separated by some experiment that depends on the mass. Positive-ray analysis at once suggests itself, but, unfortunately, quantities that are ample to give electrical or photographic evidence for the existence of an isotope correspond to minute fractions of a gram. Consequently, separation by the deflection of positive rays has been successful in only a few cases and only in minute quantities. The first methods to succeed depended on the kinetic-theory result that all the molecules in a mixture have the same average kinetic energy so that the average velocity of the lighter ones must be greater than the average velocity of the heavier ones. Thus, any experiment that depends on the average velocity of the molecules may separate those of different mass. Such an experiment is the diffusion through a small opening or system of openings. If all the molecules of the same mass had the same velocity at a given temperature, such a separation would be easy, but it will be remembered that the actual velocities of any one kind of molecule are distributed between zero and infinity. Nevertheless, a separation can be effected by various types of gas-diffusion apparatus. The ease with which such a

TABLE 17.—ISOTOPE AND ATOMIC WEIGHT TABLE OF SOME OF THE CHEMICAL ELEMENTS AS KNOWN IN JULY, 1937

1	2	3	4	5	6	7	8	9
							Chemical atomic weights, for O = 16	
Atomic number	Element	Symbol	Mass No.	Abundance, %	Packing fraction $\times 10^4$	Physical atomic weights on basis of $^{16}O = 16$	From results of isotope studies	Value from International Table
0	Neutron	n	1	....	+90	1.0090		
1	Hydrogen	H	1	99.98	+81	1.0081	1.0081	1.0078
	(Deuterium)	(D)	2	0.02	+73	2.0147		
2	Helium	He	3	....	+57	3.0171	4.0030	4.002
		He	4	100	+ 9.8	4.0039		
3	Lithium	Li	6	7.9	+27.8	6.0167	6.9374	6.940
		Li	7	92.1	+25.7	7.0180		
4	Beryllium	Be	8	0.005*	+ 9.8	8.0078		
		Be	9	99.99	+16.6	9.0149	9.0129	9.02
		Be	10	....	+16.4	10.0164		
5	Boron	B	10	20	+16.1	10.0161	10.811	10.82
		B	11	80	+11.6	11.0128		
6	Carbon	C	12	99.3	+ 3	12.0036	12.008	12.00
		C	13	0.7	+ 5.6	13.0073		
7	Nitrogen	N	14	99.62	+ 5.2	14.0073	14.008	14.008
		N	15	0.38	+ 3.2	15.0048		
8	Oxygen	O	16	99.8	± 0	16.0000		
		O	17	0.03	+ 2.7	17.0046	16.0000	16.0000
		O	18	0.16	+ 2.1	18.0038		
9	Fluorine	F	19	100	+ 2.6	19.0049	19.000	19.000
10	Neon	Ne	20	90.00	− 0.7	19.9986		
		Ne	21	0.27	......	....	20.191	20.183
		Ne	22	9.73	− 0.6	21.9985		
11	Sodium	Na	23	100	......	....	....	22.997
12	Magnesium	Mg	24	77.4	− 2.6	23.9938		
		Mg	25	11.5	......	....	24.33	24.32
		Mg	26	11.1				
13	Aluminum	Al	27	100	− 3.3	26.9911	26.985	26.97
14	Silicon	Si	28	89.6	− 5.0	27.9860		
		Si	29	6.2	− 4.7	28.9864	28.125	28.06
		Si	30	4.2	− 5.2	29.9845		
15	Phosphorus	P	31	100	− 5.0	30.9844	30.978	31.02
16	Sulphur	S	32	96*	− 5.9	31.9812		
		S	33	1*	......	....	32.05	32.06
		S	34	3*	− 5.9	33.9799		
17	Chlorine	Cl	35	76	− 5.8	34.9796	.....	35.457
		Cl	37	24	− 6.0	36.9777		
18	Argon	A	36	0.31	− 6.7	35.976		
		A	38	0.06	− 6.5	37.9753	39.953	39.944
		A	40	99.63	− 6.15	39.9754		

TABLE 17.—ISOTOPE AND ATOMIC WEIGHT TABLE OF SOME OF THE CHEMICAL ELEMENTS AS KNOWN IN JULY, 1937.—(*Continued*)

1	2	3	4	5	6	7	8	9
							Chemical atomic weights, for O = 16	
Atomic number	Element	Symbol	Mass No.	Abundance, %	Packing fraction × 10^4	Physical atomic weights on basis of $^{16}O = 16$	From results of isotope studies	Value from International Table
19	Potassium	K	39	93.4				
		K	40	0.01	Estimated	....	39.096	39.096
		K	41	6.6	as −7			
20	Calcium	Ca	40	96.76				
		Ca	42	0.77	Estimated			
		Ca	43	0.17	as −7.2	....	40.076	40.08
		Ca	44	2.30				
21	Scandium	Sc	45	100	−7	....	44.96	45.10
22	Titanium	Ti	46	8.5				
		Ti	47	7.8				
		Ti	48	71.3	−8	....	47.91	47.90
		Ti	49	5.5				
		Ti	50	6.9				
23	Vanadium	V	51	100	....	....	....	50.95
24	Chromium	Cr	50	4.9				
		Cr	52	81.6	−10	51.948		
		Cr	53	10.4	....	....	52.005	52.01
		Cr	54	3.1				
25	Manganese	Mn	55	100	....	....	....	54.93
26	Iron	Fe	54	6.5				
		Fe	56	90.2				
		Fe	57	2.8	Estimated	....	55.84	55.84
		Fe	58	0.5	as −10			
27	Cobalt	Co	57	0.17	....	....	....	58.94
		Co	59	99.83				
28	Nickel	Ni	58	67.3*	−10	57.942		
		Ni	60	26.9*				
		Ni	61	1*	....	....	58.71	58.69
		Ni	62	3.8*				
		Ni	64	1*				
29	Copper	Cu	63	70*	....	....	....	63.57
		Cu	65	30*				
30	Zinc	Zn	64	50.4	− 9.9	63.937		
		Zn	66	27.2				
		Zn	67	4.2	....	....	65.33	65.38
		Zn	68	17.8				
		Zn	70	0.4				
....	....	....	....	....	....	....	....	....

TABLE 17.—ISOTOPE AND ATOMIC WEIGHT TABLE OF SOME OF THE CHEMICAL ELEMENTS AS KNOWN IN JULY, 1937.—(*Continued*)

1	2	3	4	5	6	7	8	9
							Chemical atomic weights, for O = 16	
Atomic number	Element	Symbol	Mass No.	Abundance, %	Packing fraction $\times 10^4$	Physical atomic weights on basis of $^{16}O = 16$	From results of isotope studies	Value from International Table
35	Bromine	Br	79	50.0	− 9.0	78.929	79.91	79.916
		Br	81	50.0	− 8.6	80.926		
47	Silver	Ag	107	52.5	Estimated	....	107.87	107.880
		Ag	109	47.5	as −5.3			
53	Iodine	I	127	100	− 5.3	126.932	126.904	126.92
54	Xenon	Xe	124	0.09				
		Xe	126	0.09				
		Xe	128	1.91				
		Xe	129	26.2				
		Xe	130	4.06	....	....	131.12	131.3
		Xe	131	21.2				
		Xe	132	26.9				
		Xe	134	10.5	− 5.3	133.929		
		Xe	136	8.95				
55	Caesium	Cs	133	100	− 5	132.933	132.90	132.91
80	Mercury	Hg	196	0.10				
		Hg	198	9.89				
		Hg	199	16.45				
		Hg	200	23.77	+ 0.8	200.016	200.60	200.61
		Hg	201	13.67				
		Hg	202	29.27				
		Hg	204	6.85				
81	Thallium	Tl	203	29.4	+ 1.8	203.037	204.41	204.39
		Tl	205	70.6	+ 1.8	205.037		
82	Lead	Pb	204	1.5				
		Pb	206	28.3	Estimated	....	207.15	207.22
		Pb	207	20.1	as +1			
		Pb	208	50.1				
83	Bismuth	Bi	209	100	....	....	....	209.00
90	Thorium	Th	232	100	+ 3.0	232.070	232.02	232.12
92	Uranium	U	235	0.4	+ 3.6	235.084	238.02	238.07
		U	238	99.6	+ 3.7	238.088		

* These numbers are uncertain or tentative values.

separation can be made is obviously dependent on the relative difference of the masses of the isotopes. This is much greater in hydrogen than in other elements since one isotope has twice the mass of the other. Practically complete separation of the two isotopes of hydrogen can be achieved by gaseous diffusion.

26. But the most satisfactory method of separating the isotopes of hydrogen is by the electrolysis of water. The mechanism of this separation is not yet entirely clear and probably involves several different processes. We shall not attempt to discuss it. The experimental result is that the light hydrogen comes off at the electrodes a little faster than the heavy so that the water left behind has a slightly higher percentage of deuterium than ordinary water. By repeated electrolysis it is possible to get water that is almost pure deuterium oxide. Such water has appreciably higher density and viscosity than ordinary water and is commonly referred to as heavy water.

SUMMARY

The current in a gas discharge is carried not only by cathode rays but by positively charged ions moving in the opposite direction. Like the cathode rays, these ions can be studied by deflection in magnetic and electric fields, but they prove to be much less easily deflected. J. J. Thomson applied simultaneous electric and magnetic deflecting fields in the same direction and determined the deflections of the rays by allowing them to strike a photographic plate. The resultant deflections were at right angles. Furthermore, the traces made on the photographic plate by ions of the same m/e but different velocities were segments of parabolas. These experiments showed that the nature of the positive rays depends on the gas in the discharge, the value of m/e being approximately equal to the electronic charge divided by the average mass of the molecules of the gas. More exact results show that for any given gas there may be several different kinds of ion having different values of m/e. If this variation is assumed to be in the m, it is evident that each element may have atoms of several different masses. Such atoms, differing in mass but identical in chemical properties, are called isotopes. On the atomic weight mass scale, all atoms have approximately whole number masses even though the average mass of the atoms of an element may not be integral.

Various different combinations of electric and magnetic fields have been used by Aston, Dempster, Bainbridge, and others to study atomic masses. The exact masses of isotopes differ slightly from whole numbers. The amount of this difference is expressed by the "packing fraction." This name arises from the belief that all atoms are built out of the same few fundamental particles but that the masses of these particles are altered by the energy lost in "packing" them together. This involves the idea of the equivalence of mass and energy which comes from the theory of relativity. According to this theory, 1 kg. is equivalent to 9×10^{16} joules.

The most important isotope is the hydrogen isotope of mass two, called deuterium. Various methods have succeeded in separating isotopes.

ILLUSTRATIVE PROBLEMS

1. What is the ratio of the electric to the magnetic deflection in a Thomson apparatus in which the ratio of the magnetic to the electric field is 2 amp. turns/volt? The accelerating potential is 4,000 volts, and singly charged Be^9 ions are used.

The ratio of the electric to the magnetic deflection is x/y which from Eqs. (1) and (2), page 404, is

$$\frac{x}{y} = \frac{k\dfrac{Ee}{mv^2}}{k\mu_0\dfrac{He}{mv}} = \frac{E}{\mu_0 Hv}. \tag{6}$$

The velocity acquired by one ion due to an accelerating difference of potential of 4,000 volts is, Eq. (9), page 395, given by

$$\tfrac{1}{2}mv^2 = Ve$$

or

$$v = \sqrt{\frac{2Ve}{m}}.$$

Substituting this value of v in Eq. (6) gives

$$\frac{x}{y} = \frac{E}{\mu_0 H}\sqrt{\frac{m}{2Ve}} \tag{7}$$

where $\dfrac{E}{H} = \dfrac{1}{2}\,\dfrac{\text{volt}}{\text{amp. turns}}$ = the ratio of the electric to the magnetic field strength.

$\mu_0 = 4\pi \times 10^{-7}$ newton/amp.^{2} = the permeability of empty space.

$\frac{m}{e} = \frac{9.0149 \times 10^{-3} \text{ kg./mole}}{96{,}489 \text{ coulombs/mole}} = 9.3429 \times 10^{-8}$ kg./coulomb = the ratio of mass to charge for the Be^9 ion.

V = 4,000 volts = the accelerating potential.

The denominator in m/e is the Faraday, the charge per mole for ions carrying a single electronic charge, and the numerator the atomic weight or mass per mole. The ratio m/e may also be calculated directly though less accurately thus:

$$\frac{m}{e} = \frac{9.0149 \times 10^{-3} \text{ kg./mole}}{6.03 \times 10^{23} \text{ atoms/mole} \times 1.60 \times 10^{-19} \text{ coulomb/atom}}$$
$$= 9.34 \times 10^{-8} \text{ kg./coulomb.}$$

Substituting these values in Eq. (7) gives

$$\frac{x}{y} = \frac{1 \text{ amp.}^2 \times \text{volt}}{4\pi \times 10^{-7} \text{ newton} \times 2 \text{ amp. turns}}\sqrt{\frac{9.343 \times 10^{-8} \text{ kg./coulomb}}{2 \times 4 \times 10^3 \text{ volts}}}$$
$$= \frac{1 \text{ m.}}{2.51 \times 10^{-6} \text{ sec.}}\sqrt{1.168 \times 10^{-11} \frac{\text{kg. sec.}^2}{\text{kg. m.}^2}} = 1.36.$$

2. In Dempster's apparatus ions of identical kinetic energy are used and the accelerating potential is 100 volts. The radius of the path of the ions in the magnetic field of 2×10^5 amp. turns/m. is 4.59 cm. What is the value of m/e for these ions? If they are ions of an element carrying a single charge, what element are they? What other element carrying a double charge has approximately the same value of m/e?

As in Problem 1, the velocity of the ions is given by Eq. (9), page 395,

$$\frac{1}{2}mv^2 = Ve \qquad v = \sqrt{\frac{2Ve}{m}} \tag{8}$$

The ratio of m/e is then given by Eq. (3), page 391,

$$R = \frac{mv}{\mu_0 He}.$$

Substituting the value of v from Eq. (8) in this equation gives

$$R = \frac{m}{\mu_0 He}\sqrt{\frac{2Ve}{m}} = \frac{1}{\mu_0 H}\sqrt{\frac{2Vm}{e}}.$$

Squaring both sides, we get

$$R^2 = \frac{1}{\mu_0^2 H^2}\frac{2Vm}{e}$$
$$\frac{m}{e} = \frac{R^2\mu_0^2 H^2}{2V} \tag{9}$$

where $R = 0.0459$ m. = the radius of the path in the magnetic field.
$\mu_0 = 4\pi \times 10^{-7}$ newton/amp.^{2} = the permeability of empty space.
$H = 2 \times 10^5$ amp. turns/m. = the strength of the magnetic field.
$V = 100$ volts = the accelerating potential.

Substituting these values in Eq. (4) gives

$$\frac{m}{e} = \frac{(0.0459 \text{ m.})^2(4\pi \times 10^{-7} \text{ newton/amp.}^2)^2(2 \times 10^5 \text{ amp. turns/m.})^2}{2 \times 100 \text{ volts}}$$
$$= \frac{(4.59)^2 16\pi^2 \times 4 \times 10^{-10} \text{ newton}^2}{2 \text{ volts-amp.}^2} = 6.65 \times 10^{-7} \frac{\text{newton}^2\text{-sec.}}{\text{joule-amp.}}$$
$$= 6.65 \times 10^{-7} \frac{\text{newton-sec.}^2}{\text{m.-coulomb}} = 6.65 \times 10^{-7} \frac{\text{kg.-m.-sec.}^2}{\text{m.-sec.}^2 \text{ coulomb}}$$
$$= 6.65 \times 10^{-7} \text{ kg./coulomb.}$$

Since each ion carries a single electronic charge, the charge per mole is 96,489 coulombs. The atomic weight M is, thus,

$$M = 6.65 \times 10^{-7} \text{ kg./coulomb} \times 96{,}489 \text{ coulombs/mole}$$
$$= 6.42 \times 10^{-2} \text{ kg./mole} = 64.2 \text{ g./mole.}$$

This ion would probably be Zn^{64} since it is much more abundant than Ni^{64}. If the charge were doubled, the mass would be cut in half which would give S^{32}.

3. What is the binding energy of Li^7?

If we assume that Li^7 is made up of three hydrogen atoms and four neutrons, then the atomic weights of the parts are

$$3 \times 1.0081 + 4 \times 1.0090 = 7.0603.$$

The atomic weight of Li^7 is 7.0180 so that the energy required to break down the Li^7 atom is equivalent to $7.0603 - 7.0180 = 0.0423$ mass units. This is equivalent to

$$mc^2 = 4.23 \times 10^{-5} \text{ kg.} \times (3 \times 10^8 \text{ m./sec.})^2 = 3.81 \times 10^{12} \text{ kg. m.}^2 \text{ sec.}^2$$
$$= 3.81 \times 10^{12} \text{ joules} = \frac{3.81 \times 10^{12} \text{ joules}}{4.185 \text{ joules/cal.}} = 9.10 \times 10^{11} \text{ cal.}$$

PROBLEMS

1. Plot the magnetic deflection [Eq. (2)] against the mass of the ion for constant kinetic energy equal to 10^{-16} joule. Assume the length of the electric and magnetic fields to be 2 cm., the distance from the edge of the fields to the screen to be 20 cm., the field strength to be 10^5 amp. turns/m., and the ions to be singly charged.

2. What ratio of electric to magnetic fields in a Thomson apparatus will make the electric deflection of a 10^7 m./sec. ion equal to twice the magnetic deflection?

3. A parabola given by the equation $y^2 = 5.86 \times 10^{-3}x$ is found in Thomson's apparatus, using Cr^{52} ions. The apparatus constant k is 0.05 m.^{2},

and the magnetic field of 10^4 amp. turns/m. is equal to twice the electric field. Find the value of e/m for the Cr^{52} ions. Are the ions singly or doubly charged?

4. Suppose a stream of electrified particles each of mass m kg. and carrying a charge of q coulombs is shot horizontally with a velocity v m./sec. between two conducting plates. What must be the direction and strength of an electric field between the plates that just balances the force of gravity? What magnetic field would have the same effect?

5. In a Thomson apparatus, the ion path in the magnetic field is 10 cm. and the distance from the edge of the field to the screen is 50 cm. The magnetic field is 10^5 amp. turns/m., the electric field 2×10^3 volts/m., and e/m for Li^7 is 1.37×10^7 coulombs/kg. Find the equation of the parabola along which these ions will fall, the minimum electrostatic deflection for an accelerating potential of 5,000 volts, and the coordinates of an ion with a velocity of 2×10^5 m./sec.

6. In Bainbridge's mass spectrograph, the crossed electric and magnetic fields of the velocity selector have the values 10^4 volts/m. and 10^5 amp. turns/m., respectively. The radius of the ions in the magnetic field of 10^5 amp. turns/m. is 26.3 cm. What is m/e for these ions? If they are ions of an element carrying a single charge, what element are they? What other element carrying a double charge has approximately the same value of m/e?

7. Compute the packing fractions of D^2 and C^{13}, and compare the results with those in Table 17.

8. The relativistic kinetic energy of an electron moving with a velocity v is given by

$$m_0c^2\left[\frac{1}{\left(1 - \dfrac{v^2}{c^2}\right)^{\frac{1}{2}}} - 1\right]$$

where c is the velocity of light and m_0 the rest mass of the electron. Show that this reduces the usual form if $v << c$.

9. Using the equation of Prob. 8, calculate the difference in velocity, due to the change in mass with velocity, of the electrons in the table on page 396. Also calculate the difference for 10^6 volts.

10. Compute the increase in the mass of an electron due to a velocity of 10^8 m./sec.

11. Show that the formation of a gram molecule of helium from hydrogen would release 6.13×10^8 kg.-cal.

12. Show that 1 m. u. on the atomic weight scale is equivalent to the energy that an electron would acquire in falling through a potential difference of 931×10^6 volts. Use the relation $F = 16e/M_o$ where F is the Faraday, e the charge on the electron, and M_o the mass of the oxygen atom.

13. There is evidence that the alpha particle consists of two neutrons and two protons. On this assumption, compute the energy released in the formation of a gram molecule of helium.

CHAPTER XXI

RADIOACTIVITY

Early Experiments on Radioactivity

1. The last decade of the nineteenth century was a brilliant one for experimental physics. The discoveries of the electron, x-rays, and radioactivity followed each other in quick succession in the years 1895 to 1897. All these discoveries were closely related. Both electrons and x-rays were first observed in electrical discharge in gases at low pressure, and the discovery of radioactivity was a more or less accidental by-product of the study of certain properties of x-rays, occurring in the following way. When the pressure of the gas in a discharge tube is sufficiently reduced, the glass of the tube begins to fluoresce brilliantly and the tube begins to give off x-rays. Whether this was a chance coincidence or whether there was an intimate connection between x-rays and fluorescence was naturally an important question to the early students in this field. An interest in this question led Becquerel to study the natural fluorescence of certain compounds of uranium. He found that these compounds gave off spontaneously radiations that affected a photographic plate. Furthermore, prompted by his knowledge of x-rays, he investigated the electrical properties of these radiations and found that the air in the neighborhood of minerals containing uranium became conducting; an electroscope was rapidly discharged when a uranium mineral was brought close to it. This effect is still the simplest way of testing for the presence of radioactivity and is the one used to examine drain pipes, garbage cans, etc., when a hospital has mislaid a valuable tube of radium.

2. The initial discovery of Becquerel was quickly followed by the discovery of a whole series of substances that had radioactive properties. The activity of these substances could always be traced to the presence of one of a number of elements of very high atomic weight. Such elements were said to be "radioactive." Many of them were previously unknown, among them radium

itself, discovered by the Curies in 1898. It was found that some of these elements were very much more active than others; for example, radium is hundreds of times more effective in discharging an electroscope or in darkening a photographic plate than an equal mass of uranium. But not only do the effects of the radiations from different radioactive elements differ in degree, they differ in kind. Thus two samples of different radioactive elements that have the same effect on an electroscope when each of them is unscreened may have very different effects if they are shielded from the electroscope by a small thickness of tinfoil. In other words, the penetrating power of the radiations effective in discharging an electroscope is different for different radioactive elements. Further studies have shown that there are three kinds of radiations given off by radioactive substances. Originally differentiated only by their different powers of penetration, they were christened alpha, beta, and gamma rays in order of increasing power of penetration. These old names are still used, although we know now that the rays differ more widely than their names imply. The results of studies by more discriminating methods have given complete information concerning the nature of the rays.

The Radiations

3. Alpha rays are found to be fast-moving helium atoms that have lost two electrons and are, therefore, carrying a net positive charge of two electronic units. Both their speed and the ratio of their charge to mass can be found by electric and magnetic deflection experiments of the type described in the last chapter. The experiments are difficult, however, because the very high energies (corresponding to millions of electron volts) make it hard to produce measurable deflections. That they are really helium atoms was proved by collecting them in large numbers and showing that the material so collected had the properties of ordinary helium gas, the ions having recovered their normal quota of electrons. Alpha rays, or alpha particles as we can call them now that we know what they are, produce a great many ions along their path and are stopped comparatively quickly by material interposed in their path. In air, for example, the fastest known alpha particles penetrate for distances of 14 cm. The distance that they penetrate depends on their energy, and this in turn

depends on the particular radioactive material from which they come. The distance that they can go in air at normal temperature and pressure is called their range. Because of their enormous energies, it is possible to detect the effects of individual particles as has been described in Pars. 5 and 6 of Chap. I. Since

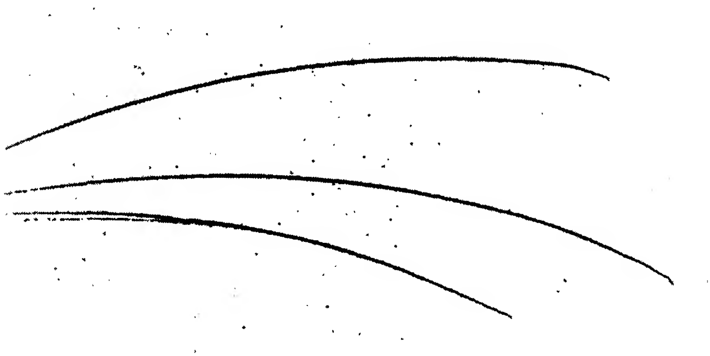


FIG. 189.—Tracks of alpha rays curved by a very intense magnetic field perpendicular to the plane of the photograph. (*Rutherford, Chadwick and Ellis.*)

their mass as well as their velocity is comparatively large, they have great momentum and are not easily deflected from the direction in which they are moving. For this reason, long sections of their paths as shown in a Wilson cloud chamber are very nearly straight lines.

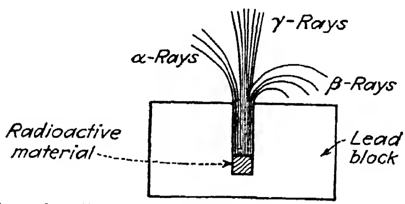


FIG. 190.—Behavior of radioactive radiations in a magnetic field directed into the plane of the figure.

4. Beta rays are found to be very fast moving electrons. Their velocities are greater than the fastest cathode rays that can be produced in the laboratory, amounting in some cases to more than 0.9 the velocity of light. In spite of their great speed, the same methods used for the study of cathode rays can be applied to

them. They produce less intense ionization than the alpha particles but are more penetrating and more easily deflected by collisions with atoms along their path. The ionization they produce is still sufficiently strong to produce visible paths in a

FIG. 191.—Beta-ray tracks. (*Rutherford, Chadwick and Ellis.*)

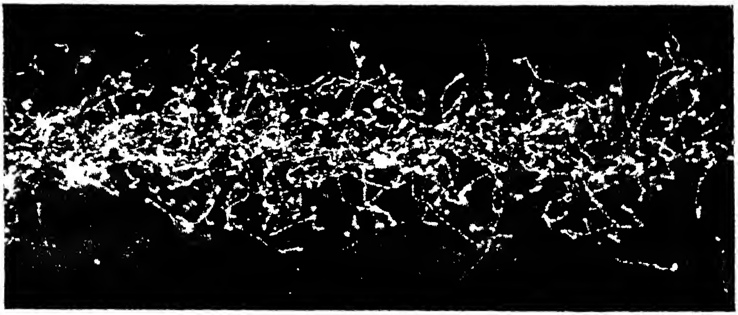


FIG. 192.—Tracks of secondary electrons produced by a beam of x-rays passing from left to right. (*C. T. R. Wilson.*)

Wilson cloud chamber, but these tracks are much weaker than those of the alpha particles.

5. Gamma rays are the most penetrating of all the radioactive radiations and are not affected at all by either electric or magnetic fields. It is found that they are not particles but electromagnetic

radiations like x-rays, light or radio waves. They are of exceedingly small wave length, of the order of 10^{-12} m.

Radioactivity and the Structure of Matter

6. It is surprising that the phenomena of radioactivity furnish information about the structure of matter from two entirely different points of view. One point of view ignores the parent substances and concerns itself only with the passage of the radiations through matter. It simply considers the radiations, particularly the alpha and beta radiations, as tools providentially supplied by nature and uses these tools as probes with which to pry into the inner structure of atoms. The laboratory cannot, or at least until the last few years could not, provide tools of comparable power. It is this field of study that we shall discuss first. The other point of view, which concerns itself primarily with the radioactive elements themselves, will be taken up later on in the chapter.

The Scattering of Alpha Particles

7. In the kinetic theory of gases, we talked a good deal about the collisions between molecules and gave an estimate of the size of the molecules and the distances they travel between collisions. If these results are applied to the passage of alpha particles through a gas, the long straight paths observed in the Wilson cloud chamber become inexplicable. According to the kinetic theory, the alpha particles must make many hundreds of collisions for every millimeter that they travel through the air in the chamber. Apparently a charged atom of helium of high velocity behaves very differently from an uncharged atom moving with ordinary thermal velocities (alpha-particle kinetic energies are of the order of 10^9 times as large as thermal kinetic energies). Evidently what would be collisions for a slow-moving atom are hardly noticed by the fast moving alpha particle. This can be understood only if the atoms of the material through which the alpha particle is moving are not the solid elastic billiard-ball structures assumed by kinetic theory but have a structure through which an atomic projectile can break with comparative ease if it is moving fast enough. We notice, however, that every once in a while an alpha-ray track shows a sharp bend. It does occasionally hit something in an atom that is solid enough to deflect it, but

apparently this obstacle is much smaller than the atom as a whole, since the alpha particle passes through many atoms without being deflected.

8. The experiments on the scattering of alpha particles are so important and so essentially simple that we shall give a brief description of one of the methods used in Rutherford's laboratory. A schematic drawing of the apparatus is given in Fig. 194. The

FIG. 193.—Deflections of an alpha particle by collisions near the end of its path. (*Rutherford, Chadwick and Ellis.*)

radioactive source of alpha particles is put at the back of a hole in a lead block. Limited in direction by the screening effect of the lead, the alpha particles emerge in a beam to strike the scattering target of thin metallic foil. Most of the particles go through the foil, coming out at various angles. Some of them then strike the zinc sulphide screen, making scintillations that can be observed through the microscope. The box enclosing the whole apparatus

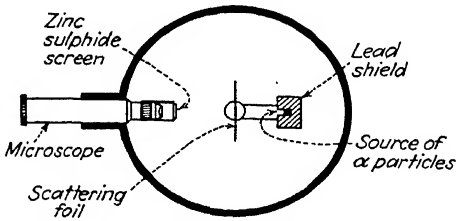


FIG. 194.—Schematic diagram of apparatus for study of alpha-particle scattering.

is airtight and can be evacuated. Furthermore, it is mounted on a ground joint so that it can be rotated about an axis perpendicular to the plane of the drawing and passing through the center of the foil. Since the microscope and zinc sulphide screen are mounted on the box, they rotate with it, whereas the source and scattering foil remain fixed. Alpha particles from the source strike the foil. It is so thin that many of them make no "collisions," and the chance of one of them making more than one

collision is negligibly small. Those which go straight through strike the screen and make scintillations that can be counted by the observer at the microscope. Those which make collisions are "scattered" through various angles. The number scattered through different angles can be counted by turning the screen and microscope to different positions and observing the scintillations. Early experiments showed a comparatively small number of scattered particles, but a surprisingly large proportion of those deflected at all was deflected through large angles. To explain this result, Rutherford developed a theory based on definite assumptions concerning the structure of atoms. Not only was his theory of alpha-particle scattering completely confirmed by later more exact observations, but his views on atomic structure have been triumphantly successful in other fields. They are the foundation on which the whole current theory of atomic structure is based.

The Rutherford Atom

9. We have seen that all atoms contain electrons and something else that is much heavier and carries enough positive charge to counteract the negative charge of the electrons. According to Rutherford, this positive charge is concentrated at the center of the atom where most of the mass is also concentrated. This positively charged heavy nucleus is very small, of the order of 10^{-14} m. in diameter, about one ten-thousandth of the diameter of the atom obtained from kinetic theory. Around the nucleus, in a sort of planetary system, enough electrons are arranged to counteract the charge on the nucleus. The region not occupied by either electrons or nucleus is empty space. When two atoms with thermal kinetic energies come together, the electric forces of repulsion between the outer electronic structures are sufficient to drive the atoms apart so that, in such a "collision," the centers of the atoms never come closer together than the diameter of the outer electronic structures. But when an alpha particle of three or four million volts kinetic energy comes along, it brushes aside the small masses of the electrons and continues with its momentum practically unaltered unless it happens to get very close to the nucleus. The cumulative effect of many millions of collisions with electrons is to use up all of the kinetic energy of the alpha particle and bring it to rest at the end of its range. The loss of

energy at any one such collision is, however, a negligible fraction of the total energy; to be specific it amounts, on an average, to about 30 electron volts per collision.

10. If, however, the alpha particle does come close to the nucleus, the ordinary inverse-square law of repulsion between electric charges comes into play and the particle and the nucleus repel each other. In general, both alpha particle and nucleus move, but if the nucleus is much the heavier of the two as is usually the case, its motion will be slight and the resultant effect is a deflection of the alpha particle. The theory of this deflection involves only the laws of momentum and electricity with which we are familiar, but it is somewhat complicated as to geometry and algebra so will not be given here. Suffice it to say that it involves the amount of the positive charge on the nucleus and that experiments can be done with sufficient precision actually to determine this charge in a few instances. It is found to be different for different elements and always to be equal to the atomic number of the element in question. The atomic number of an element is its position in the periodic table, if we count from left to right and down beginning with hydrogen. The atomic numbers are given in the periodic table at the back of the book. This is, of course, an empirical definition. Now that we have considered the results of the study of alpha-particle scattering, we would do better to say that

The atomic number of an element is the number of positive charges on the nucleus of an atom of that element. The atomic number of an element determines its chemical properties and, therefore, its position in the periodic table.

Radioactive Disintegration

11. We have been focusing our attention on the radiations from radioactive substances and the way in which they can be used to study the structure of matter. We now wish to turn to the substances themselves. Here again, we shall learn much about the way in which atoms are built but from a different point of view. It is found that both the nature and intensity of the radioactivity of a given substance vary with the time. It appears, therefore, that a radioactive substance changes its character as the result of its activity. In the light of our knowledge of the nature of the radiations and of the nature of the

atomic nucleus according to Rutherford's hypothesis, we can interpret this change if we assume that the radioactive radiations come from the nucleus of the atom. To be specific, suppose we take radium itself as an example. Radium is chemically similar to barium, has an atomic number of 88 and an atomic weight of approximately 226. It emits alpha particles. If they come from the nucleus, the charge of the nucleus must be reduced by two units and its mass reduced by four units by their emission since each alpha particle is a helium nucleus of mass four and carrying two positive charges. Each radium atom that loses an

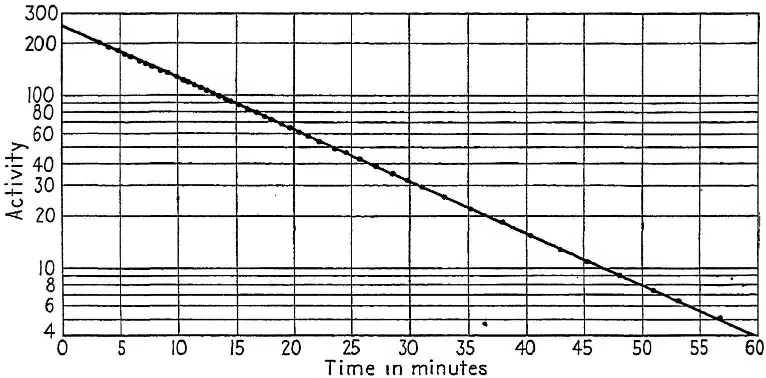


FIG. 195.—Electroscope readings on radioactivity of copper plotted logarithmically against time. Radioactive copper had been formed by alpha-particle bombardment of cobalt. (*Ridenour and Henderson.*)

alpha particle must therefore change to an atom of an element having an atomic number of 86 and an atomic weight of approximately 222. But an element of atomic number 86 would lie in the same column of the periodic table as the rare gases xenon, krypton, argon, neon, and helium. We might expect, therefore, that radium would give off a gas having properties similar to those of the above-named gases, *i.e.*, that was chemically inert. This is found to be true. It is not difficult to separate an inert gas from a barium-like metal so that it is possible to get a radium preparation that is momentarily free of its emanation. It is then possible to study the rate at which the emanation is produced. It is found that the emanation is produced as if there were a certain chance that any given atom of radium would throw out an alpha

particle at any time. Or speaking statistically, if we have a large number of atoms of radium present, a certain fraction of them will disintegrate in a certain time and the number disintegrating in an infinitesimal time interval is proportional to the number present at the beginning of the interval. In mathematical language, $dN/dt = -\lambda N$ where N is the number of atoms present, dN is the number disintegrating in the time dt, and λ is a constant of proportion characteristic of the particular radioactive substance concerned. By integral calculus, it is easy to proceed from this equation giving the instantaneous rate of change of the number of atoms present to a second one $N = N_0e^{-\lambda t}$ which gives N, the number of atoms left after a time t in terms of λ, t, and N_0 where N_0 is the number of atoms of the radioactive atoms present originally.* This law is found to be a general one applying to all radioactive disintegrations; a result that is not surprising since we are really dealing merely with the law of pure chance. There is, however, a very great difference in the rate at which different radioactive disintegrations proceed.

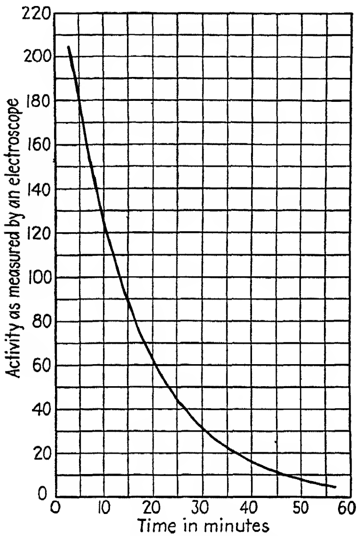


FIG. 196.—Exponential decay curve for artificial radioactive copper. Experimental points were originally plotted on a logarithmic scale as shown in Fig. 195. This curve is derived from that.

According to our mathematical formulation of the law of decay, the rate at which the number of atoms N decrease depends upon the value of the constant λ. Substances decaying at different rates will accordingly be characterized by different values of λ. An equivalent way of describing the difference in rates of decay is to specify the so-called half life of the elements. This is the time during which half of the atoms originally present will have disintegrated. It can be shown to be equal to $0.693/\lambda$. The half

* e is the base of the Napierian or natural system of logarithms.

lives of different radioactive elements vary from many thousands of years to a small fraction of a second.

12. The change of atomic number and atomic weight cited in the last paragraph for radium is characteristic of any alpha-particle emission. The corresponding transition in the case of the emission of a beta particle is an increase of one in the atomic number since the removal of a negative charge increases the net positive charge on the nucleus. The new element produced by such a transition would therefore be in the column of the periodic table to the right of that containing the parent element. The

TABLE 18.—RADIOACTIVE DATA

Element uranium series	Symbol	At. No.	At. wt.	Half life	Radiation	Range in air, cm.	Velocity in fractions of speed of light
Uranium I.......	UI	92	238	4.5×10^9 yr.	α	2.37	0.0456*
Uranium X_1.....	U-X_1	90	234	23.8 days	β		
Uranium X_2.....	U-X_2	91	234	1.15 min.	$\beta(\gamma)$		
Uranium II......	UII	92	234	2×10^6 yr.	α	2.75	0.0479*
Ionium..........	Io	90	230	6.9×10^4 yr.	α	2.85	0.0485*
Radium.........	Ra	88	226	1,580 yr.	$\alpha(\beta, \gamma)$	3.13	0.0500*
Radon..........	Rn	86	222	3.85 days	α	3.94	0.054*
Radium A.......	Ra-A	84	218	3.0 min.	α	4.50	0.0565*
Radium B.......	Ra-B	82	214	26.8 min.	$\beta(\gamma)$	....	0.74* (max.)
Radium C.......	Ra-C	83	214	19.5 min.	99.97 % β, γ	....	0.96* (max.)
Radium C'....	Ra-C'	84	214	10^{-6} sec.	α	6.57	0.0641*
Radium D (Radiolead)..	Ra-D	82	210	16.5 yr.	β, γ	....	0.39* (max.)
Radium E.....	Ra-E	83	210	5.0 days	β		
Radium F (Polonium)..	Ra-F	84	210	136 days	$\alpha(\gamma)$	3.58	0.0523*
Radium G.....	Ra-G	82	206				
Radium C.......	Ra-C	83	214	....	0.03 % α		
Radium C''....	Ra-C''	81	210	1.4 min.	β		

* To get velocity in meters per second multiply by 3×10^8.

For similar data on the actinium series and the thorium series see the International Table of the Radioactive Elements and Their Constants in the "Handbook of Chemistry and Physics" of the Chemical Rubber Company.

mass of the ejected beta particle is too small to affect the atomic weight noticeably. The emission of gamma rays does not affect the atomic number at all or the atomic weight to an appreciable extent. In fact, the gamma rays are a secondary consequence of the rearrangement of the atomic nucleus after the emission of an alpha or beta particle.

The Displacement Laws

13. These laws describing the change of the character of an element, resulting from the emission of an alpha or beta particle, are called the displacement laws and may be stated as follows:

The emission of alpha particles by an element results in the production of an element two places lower down in the natural sequence of the elements and having an atomic weight four units less than that of the parent.

The emission of beta particles by an element produces an element one place higher in the sequence of the elements and having the same atomic weight as the parent.

The Radioactive Series

14. By a combination of the electrical methods appropriate to the study of the radioactive radiations and chemical methods

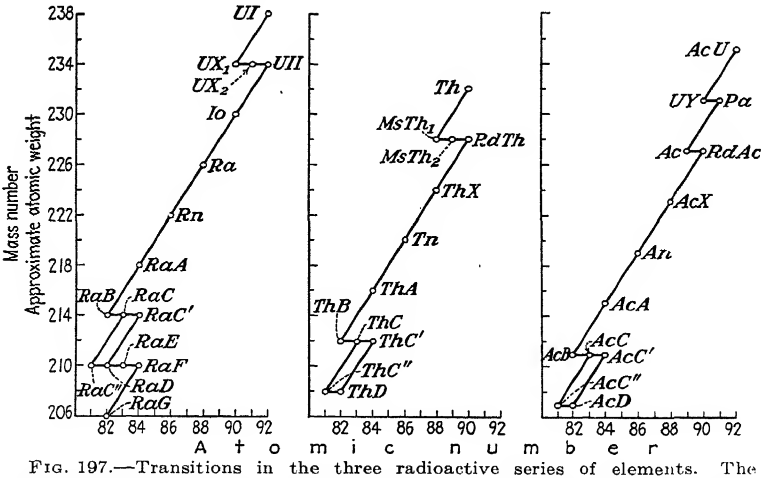


FIG. 197.—Transitions in the three radioactive series of elements. The parent element at the top of each series becomes various other elements by successive alpha- and beta-particle transitions ending at lead. The half lives are not given. Those for the radium series can be found in Table 18, page 437.

appropriate to the determination of the combining properties and atomic weights of the radioactive elements, it has been shown that there are in the neighborhood of forty kinds of radioactive atoms found in nature. They are divided into three groups or series each derived by radioactive transformation from a single

parent element of long half life. These three series are known as the uranium, thorium, and actinium series. In each series, the parent element is one of high atomic weight and number that disintegrates by a series of alpha and beta transformations through elements of varying life times and chemical properties until a nuclear structure is reached which is stable. The successive steps in these transformations are shown in Fig. 197. The first two steps in the actinium series have only recently been established, and there are still some unsettled points about the details of all three series but the main features of them are fully established. The diagram has been drawn to bring out the significance of the "displacement laws" already discussed, namely, the changes that occur in atomic number and atomic weight as a result of the emission of alpha and beta particles from the nuclei. A discussion of the variations in the half lives of the different elements is somewhat beyond the scope of this book. We can simply say in passing that they are connected with the relative stability of the nuclei and that this also determines the energy of the emitted particle. In general, the shorter the half life, the greater the kinetic energy of the ejected particle.

Isotopes

15. Comparison of the three radioactive series shown in Fig. 197 shows that they are strikingly similar. If we compare the atomic numbers of the elements in the three series, we find that there are numerous cases of elements of identical atomic number which have yet come from different parents, have different atomic weights and different offspring. Furthermore, such elements have different half lives. Earlier in this chapter it was stated that the chemical properties of an element were determined by the atomic number. Accordingly, we would expect that any two of the radioactive elements that have the same atomic number would be chemically identical, and this is found experimentally to be true. For example, the element radon of atomic number 86 and atomic weight about 222, which has already been discussed, finds analogues in both the thorium and actinium series at the same atomic number but different atomic weights; these other two elements are also inert gases and chemically inseparable from radon. Or consider the end products of all three series. They all have the atomic number 82, and consultation of the periodic

table shows that this is the atomic number of lead. The end products of the radioactive series are all found to be in fact chemically identical with lead. According to these views, the atomic weight of lead that was predominantly made up of atoms recently formed by radioactive decay might differ from the atomic weight of ordinary lead and, furthermore, might differ according to which radioactive series had given rise to it. Careful determinations of the atomic weight of lead from various radioactive ores have confirmed this expectation.

16. The discovery of atoms of identical chemical properties but differing in mass and radioactive properties was of fundamental importance to our understanding of the structure of matter. It was this radioactive work that first suggested the existence of isotopes whose presence in the nonradioactive elements we discussed in the preceding chapter. As we saw there, it has been decided to call all atoms by the name of the element identified with their chemical properties even though their masses and radioactive properties may be different. Thus we still say that we have only 92 different chemical elements although we know that we have a good many more than that number of different kinds or species of atoms.

Artificial Radioactivity

17. In the last few years, the subject of radioactivity has taken on new life. Previously, laboratory methods had not succeeded in affecting the natural processes of radioactive disintegration in any way. The life times of the radioactive elements, the energies of the radiations they emitted, etc., were entirely unaffected by anything that could be done to them. Nor had it been possible to disrupt the nucleus of the lighter elements except in a very small way by using the high-energy alpha particles from natural radioactive substances. A gradual improvement in laboratory technique coupled with a better theoretical understanding of the problems involved suddenly began to bear fruit about 1932. The results obtained since that time have been amazing. It is now possible to change one nucleus into another and to produce nuclei unknown in nature. Such nuclei are themselves radioactive so that such machines as the cyclotron are now able to produce artificial radioactive substances. We shall return to these topics in a later chapter (XXVII). We can hardly hope

to discuss them intelligibly until we have covered the subject of electromagnetic radiation.

SUMMARY

In 1896, Becquerel found that certain uranium minerals will discharge an electroscope in their neighborhood. This is a property characteristic of three different groups of elements at the high-atomic-weight end of the periodic table. Such elements are called radioactive and produce ionization in air or other gases because they emit three different types of radiations. Rays of the first type, the least penetrating, are called alpha rays. They are helium atoms carrying a double positive electronic charge and moving with high velocities. Rays of the second type, more penetrating, are called beta rays and are very high speed electrons. Rays of the third type, gamma rays, are not particles at all but electromagnetic waves of very short wave length. They are extremely penetrating.

Radioactivity gives information about the structure of atoms in two ways. In the first place, the rays are used as high-speed projectiles. Alpha rays travel for considerable distances through apparently solid matter with a few deflections through large angles. Rutherford interpreted these results as showing that an atom consists of a very small, very heavy, positively charged nucleus surrounded by a system of electrons at comparatively large distances. The positive charge on the nucleus is the atomic number of the atom and determines the number of electrons around the nucleus and, therefore, its chemical properties.

The changes that occur in radioactive elements when they send out radiations are the second source of information that radioactivity provides. A given element emits either alpha particles or beta particles but not both. The rate of emission and the energy vary from element to element. The emission changes the chemical nature of the element, showing that the particles come from the nucleus. The emission of an alpha particle reduces the atomic number by two and the mass number by four. The emission of a beta particle increases the atomic number by one and does not affect the mass number. There are three series of radioactive elements in which one element after another is formed by successive disintegration. In recent years, artificial radioactive elements have been made.

ILLUSTRATIVE PROBLEMS

1. If an alpha particle from radium makes a perfectly elastic head-on collision with a Hg^{200} nucleus at rest, what will be the final velocity of particle and nucleus?

The velocity of an alpha particle from radium is 1.51×10^7 m./sec. Its mass is 4.039 mass units. The mass of Hg^{200} is 200.016 mass units. Since the velocity of the mercury atom is initially zero, we may use Eqs. (7) and (8) of page 116,

$$v_a = \frac{m_a - m_b}{m_a + m_b} u_a \qquad v_b = \frac{2m_a}{m_a + m_b} u_a \tag{1}$$

where v_a = the final velocity of the alpha particle.
v_b = the final velocity of the Hg^{200} atom.
$u_a = 1.51 \times 10^7$ m./sec. = the initial velocity of the alpha particle.
$m_a = 4.039$ m. u. = the mass of the alpha particle.
$m_b = 200.016$ m. u. = the mass of the Hg^{200} atom.

Substituting these values in Eq. (1) gives

$$v_a = \frac{4.039 \text{ m. u.} - 200.016 \text{ m. u.}}{4.039 \text{ m. u.} + 200.016 \text{ m. u.}} 1.51 \times 10^7 \text{ m./sec.} = -1.45 \times 10^7 \text{ m./sec.}$$

$$v_b = \frac{2 \times 4.039 \text{ m. u.}}{4.039 \text{ m. u.} + 200.016 \text{ m. u.}} 1.51 \times 10^7 \text{ m./sec.} = 5.98 \times 10^5 \text{ m./sec.}$$

The minus sign indicates that the alpha particle reverses its direction.

2. If half the atoms of Ra D disintegrate in 16.5 yr., how many atoms in 2 mg. of Ra D will disintegrate in 1 yr.?

The radioactive disintegration constant

$$\lambda = 0.693/T = 0.693/16.5 \text{ yr.} = 0.0420 \text{ yr.}^{-1}.$$

The number of atoms in 1 g. of Ra D is Avogadro's number divided by the atomic weight in grams, or $\dfrac{6.03 \times 10^{23} \text{ atoms/mole}}{210 \text{ g./mole}} = 2.87 \times 10^{21}$ atoms/g. or 2.87×10^{18} atoms/mg. There are thus

$$2 \times 2.87 \times 10^{18} \text{ atoms} = 5.74 \times 10^{18} \text{ atoms in 2 mg. of Ra D.}$$

The number of atoms left after 1 yr. is given by $N = N_0e^{-\lambda t}$ from Par. 11. The number of atoms that disintegrate in 1 yr. is therefore

$$N_0 - N = N_0(1 - e^{-\lambda t}) = 5.74 \times 10^{18} \text{ atoms } (1 - e^{-0.0420 \text{ yr.}^{-1} \times 1 \text{ yr.}})$$

$$e^{-0.0420} = \frac{1}{e^{0.0420}}, \qquad \log_e e^{0.0420} = 0.0420,$$

$$e^{0.0420} = 1.043, \qquad e^{-0.0420} = \frac{1}{1.043} = 0.959$$

$$N_0 - N = 5.74 \times 10^{18} \text{ atoms } (1 - 0.959) = 2.4 \times 10^{17} \text{ atoms.}$$

PROBLEMS

1. How much would the loss of 30 electron volts by collision with an electron slow down an alpha particle having a velocity of 1.51×10^7 m./sec.?

2. How many collisions with electrons must an alpha particle from radium make to be stopped?

3. If an alpha particle from Ra C′ (velocity 1.92×10^7 m./sec.) makes a perfectly elastic head-on collision with an Au nucleus at rest, what will be the final velocity of alpha particle and nucleus?

4. Show that the half life of a radioactive element is equal to $0.693/\lambda$.

5. How long a time will it take for all the atoms of a radioactive substance to disintegrate?

6. What is the mass of the electron on the atomic weight scale?

7. If the difference of the reciprocal half lives of two elements is two per second, what is the ratio of the fractions disintegrating each second?

8. If the ratio of the half lives of two elements is 3:1, show that $F_1 = (F_2)^3$ where F_1 and F_2 are the fractions of the longer and shorter lived elements respectively that will disintegrate in a given time.

9. If half the atoms of radium disintegrate in 1,580 years, find the fraction of atoms that will disintegrate in one day.

10. If the half life of radium is 1,580 years, how many atoms in 1 mg. of radium will disintegrate in 1 sec.?

CHAPTER XXII

ELECTROMAGNETIC INDUCTION

1. We have seen that electrons in conductors can move about with considerable freedom, and we have also seen that electrons moving in a magnetic field are acted on by a force μHev perpendicular to their direction of motion and to the magnetic field. We have good reason to suppose that this force acts on electrons whether they are in a cathode-ray beam, inside a conductor, or anywhere else. Consequently, the electrons in a conductor which is moving in a magnetic field will be subjected to this force and a current will start to flow. This is the principle of electromagnetic induction on which the working of all dynamos is based.

2. Let us consider a simple specific case. Suppose a single wire hanging between the poles of an electromagnet is steadily moved sideways with a uniform velocity v perpendicular to the magnetic field. Every electron in the wire will experience a force μHev in a direction along the wire. If we are dealing with a short straight piece of wire such as is shown in Fig. 198, electrons will move in the direction shown until there is an accumulation at A and a deficit at B; in other words, until there is a negative charge at A and a positive charge at B. These charges will exert electrostatic repulsion and attraction on the electrons in the wire tending to move them toward B, *i.e.*, the electrostatic forces are in the opposite direction to the forces set up by the motion in the magnetic field. As more electrons move from B to A, the electrostatic forces increase, retarding the motion until it finally ceases entirely. This whole process is similar to the induction of static charges that would occur if the positive terminal of a battery or static machine were brought up to A and the negative terminal to B without actually making contact (Fig. 199). Thus

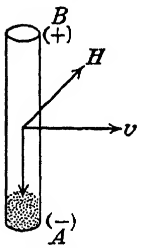


FIG. 198.—The motion of a wire with a velocity, v, across a magnetic field, H, causes some of the electrons in the wire to accumulate at A.

the effect of moving the wire in a magnetic field perpendicular to it is the same as the effect of having a stationary electric field parallel to it. The force on an electron in such an electric field would be Ee if E were the strength of the field and e the charge on the electron. But we have seen that the force on every electron in the wire resulting from its motion in the magnetic field is μHev; therefore, the effect of this motion is equivalent to an electric field of strength $E = \mu Hv$.* If the motion is not perpendicular to the magnetic field but at an angle θ to it, only the component of motion perpendicular to the field is effective. Hence, the general law is that the induced force per unit charge is equal to $\mu Hv \sin \theta$, is perpendicular to the plane determined by μH and v, and is in the direction such that a right-hand rotation of θ deg. around it would turn v into H.

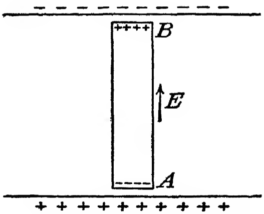


FIG. 199.—The presence of an electric field, E, set up by two oppositely charged plates, produces the same effect as the motion in the magnetic field of Fig. 198.

3. The elementary problems which we shall consider are more easily treated in terms of induced e.m.f. instead of induced electric field strength. Consider the piece of wire of Par. 2 as part of a complete electric circuit, the rest of which is outside the magnetic field. Then the flow of electrons will be continuous without the accumulation of charges at any point. The force on the electrons in the section AB of the circuit causes current to flow just as if a battery were present in the circuit. The e.m.f. of this equivalent battery may be called V, and it is simply equal to the difference of potential between the points A and B of the wire. If L is the length of the wire, V/L is the electric field strength at every point between A and B and the force on a charge e in the wire will be Ve/L. But we have already seen that this force is μHev so that we have $V/L = \mu Hv$ or $V = \mu HvL$ for the induced e.m.f. Perhaps this can be seen more clearly from the point of view of work. By definition, the potential difference between

* We do not need to say then that moving a wire perpendicular to a magnetic field of magnitude H with a velocity v actually induces in the wire an electric field $E = \mu Hv$, but, as far as the force on the electrons in the wire is concerned, motion in a magnetic field is equivalent to the presence of an electric field.

the points A and B is the work required to move unit charge from B to A. But the force on such a charge will be $F = \mu Hv$ (since the charge is one); therefore, the work required is $FL = \mu HvL$, and this is the potential difference between A and B. This is the same result we just obtained from a slightly different point of view. A potential difference set up in this way by motion in a magnetic field is called an *induced electromotive force*. We have shown how it results from the motion of a conductor in a magnetic field, but from the point of view of the electron inside the conductor it is immaterial whether it is the wire or the field that is moving. It is merely necessary that the magnetic field and the wire be moving relative to each other. However, to understand the situation in the most general way, we need to express our result in a new form.

E.M.F. Induced in a Closed Circuit

4. Let us now consider the straight wire already discussed as one side of a rectangle of wire set perpendicular to a uniform magnetic field. Suppose that this side AB (Fig. 200) can slide along parallel to itself maintaining electrical contact with the sides AD and BC. If the wire has a length L and is moving with a velocity v, then the e.m.f. induced in it is μHvL, as we have seen, and this is the total e.m.f. induced in the circuit if the other parts are not moving and the magnetic field is constant. This induced e.m.f. will be given in volts if the other quantities are in m.k.s. units.

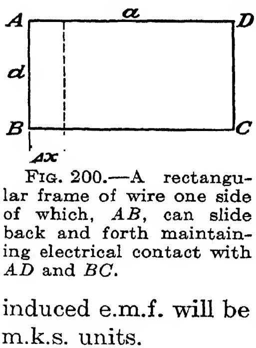


FIG. 200.—A rectangular frame of wire one side of which, AB, can slide back and forth maintaining electrical contact with AD and BC.

5. Now consider the product μHS where S is the area of the circuit perpendicular to the magnetic field. In a small time interval Δt, the area of the circuit will be changed by the motion of the sliding wire. The wire moves a distance $\Delta x = v\Delta t$, and therefore the area S is diminished by $\Delta S = L\Delta x = Lv\Delta t$. Therefore, the product μHS is changed by $\mu H\Delta S$ in the time Δt, and its rate of change is

$$\frac{\mu H\Delta S}{\Delta t} = \mu HLv. \tag{1}$$

But this we have just seen to be equal to the e.m.f. induced in the circuit; so we have proved that the induced e.m.f. is equal to the rate of change of the product μHS.

Magnetic Flux

6. Because the result of the last paragraph proves to be a perfectly general one, we shall restate it in different terms. First of all, we shall give a name to the product μHS. We shall call this product the magnetic flux through a circuit and define it in a general way as follows:

The magnetic flux through any circuit is the product of the surface area enclosed by the circuit, the component of the magnetic field strength perpendicular to the surface, and the permeability of the material in which the circuit is placed (Fig. 201). If the surface is not plane and the field not uniform, the flux is obtained by breaking up the surface into small bits over which the field can be considered uniform and the surface plane, and then summing up the contributions of all these small elements, in other words, by a process of integration. The magnetic flux is usually designated by the Greek letter Φ, and therefore we can write our definition in mathematical language as

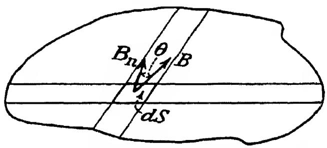


Fig.—201.—Magnetic flux through a circuit. The flux through any element of area dS is $B_n dS$ where $B_n = B \cos \theta$ is the component of magnetic induction normal (perpendicular) to dS. The total flux through the circuit is the sum of the contributions from all the elements of area, dS.

$$\Phi = \mu H_n S. \tag{2}$$

7. We should recall that we defined the magnetic induction B as the product μH so that we can write the preceding equation as

$$\Phi = B_n S \tag{3}$$

and define the magnetic flux as the product of the normal component of the magnetic induction and the area. Or, as is sometimes more useful, this can be reversed and the magnetic induction defined as the magnetic flux per unit area.

Faraday's Law of Induction

8. Comparing the definition of magnetic flux that we have just given with the result we obtained for the e.m.f. induced in a circuit, we see that the induced e.m.f. is equal to the rate of change of the magnetic flux. Furthermore, if we examine the direction of the induced e.m.f. in AB according to the motor law as in the first paragraph, we see that if the flux is diminishing the direction of the induced current is such as to maintain the flux, and vice versa. Both these results are perfectly general and are expressed in Faraday's Law of Induction as follows:

The electromotive force induced in a closed circuit of one turn is equal to the rate of change of magnetic flux through that circuit and is in such a direction that the induced current opposes the change in magnetic flux.

In mathematical language, by using the calculus notation where d/dt denotes rate of change with time,

$$V = \frac{d\Phi}{dt}. \tag{4}$$

Evidently if the flux is changing at a constant rate, the induced e.m.f. is constant and the change in flux is equal to the product of the induced e.m.f. and the time, *i.e.*, $\Phi = Vt$; and since V is in volts and t in seconds, the unit of magnetic flux in the m.k.s. system is the volt-second. This unit of magnetic flux is called the weber. Evidently the unit of magnetic induction is the volt-second per square meter or the weber per square meter. We can now get an idea of the physical meaning of the flux. If the flux disappears entirely in 1 sec., the average e.m.f. in the circuit during that time is just numerically equal to the flux.

9. We have derived the law of induction for the special case of a rectangular circuit where the magnetic flux is changed by moving the wire on one side of the rectangle. In this case, it was easy to relate the result to the earlier one for the force on a charge moving in a magnetic field. But the result is more general than such a relation implies. Even when no simple interpretation can be given in terms of electrons moving in a magnetic field, there is found to be a resultant e.m.f. induced whenever the magnetic flux through a circuit is changing. In this book, we must take this merely as an experimental fact. Actually, a theoretical

relation to the other interpretation does exist, though it is an extremely subtle one.

Lenz's Law

10. In the simple case of the motion of a conductor through a magnetic field, it was easy to find the direction of the induced e.m.f. by applying the rule already given for the force on a charge moving in a magnetic field; but in the statement of Faraday's law of induction, we included a more general principle that covers all cases of induced currents. This principle is, in fact, the electrical analogue of Newton's first law in mechanics, suggesting that nature dislikes change in electrical conditions as in conditions of motion. It is called Lenz's Law and is of sufficient importance to repeat separately as follows: An induced current always flows in a direction such that its magnetic field opposes the change in magnetic flux inducing the current.

11. For example, in the rectangular circuit discussed in Par. 3, if the wire AB is moved in a direction to reduce the area of the circuit so that the magnetic flux is diminishing, according to Lenz's law the induced current will flow in a direction to increase the flux. This means that if the original magnetic field is into the paper, the induced current must flow clockwise so that the field it sets up is in the same direction. If, on the other hand, the wire AB is moved in the other direction, tending to increase the flux. the induced current flows counterclockwise so that its field is opposite to the original field. These directions are the same as those derived by applying the "motor law" as in Par. 2. It is to be noted that the induced current flows in a direction that not only opposes the change of flux, but also is the direction such that the motor force of the magnetic field on the induced current opposes the motion producing the change in flux.

Some Examples of Induced Currents

12. In Par. 4, we considered a rectangular circuit with one side that could be moved. Let us now consider a rigid rectangular conductor under various conditions of motion in magnetic fields.

13. *Translational Motion in a Uniform Field* (Fig. 202). If the plane of the conductor is perpendicular to the field and the motion is perpendicular to AB, one side of the rectangle, there

must be an induced e.m.f. in that side; but if the field is uniform, there must be an equal and opposite induced e.m.f. in CD, the opposite side. The induced electric fields in BC and DA are across the conductor and, therefore, make no contribution to the resultant e.m.f. driving current around the circuit. The resultant induced e.m.f., therefore, is zero. This agrees with the point of view of change of flux since there will be no change of flux if the circuit is moving in a uniform field and not changing its orientation with respect to it.

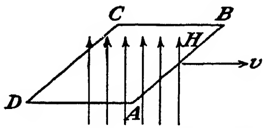


FIG. 202.—A rectangular circuit moving across a uniform magnetic field. No resultant induced e.m.f.

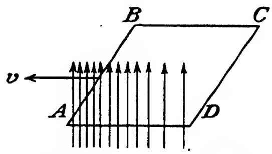


FIG. 203.—A rectangular circuit moving across a non-uniform magnetic field. Changing flux induces an e.m.f.

14. *Motion in a Nonuniform Field.* If the side AB of the rectangle is moving in a region of more intense magnetic field than the side CD, then the e.m.f. induced in it will be greater than the opposing one in CD so that there will be a resultant induced e.m.f. and current. In this case, the flux through the circuit is clearly increasing. Referring to Fig. 203, we see that according to the "motor" law the force on a positive charge in AB will be from A to B and in CD from D to C but smaller; therefore, the resultant induced current flows in the direction $ABCD$. Let us consider whether this is in accord with Lenz's Law as stated above. According to it, the induced current should be in a direction to oppose change. Therefore, it should be in a direction to reduce the resultant magnetic flux through the circuit since the motion is tending to increase it. This means that the magnetic field of the induced current should point downward inside the rectangle since the inducing field is upward. If we apply the right-hand rule to the current, we see that it must flow in the direction $ABCD$ to fulfill this condition.

15. Another way to apply Lenz's Law is to the motion itself. As soon as the induced current begins to flow, there will be a force on the sides of the rectangle because they are carrying current in a

magnetic field. According to Lenz's Law, this force must be in a direction to oppose the motion, *i.e.*, the resultant force from this cause should be from left to right in Fig. 203. If the current is in the direction *ABCD*, the force on *AB* is from left to right, on *BC* is toward *AD*, on *CD* is from right to left, and on *DA* is toward *BC*. If, as we have been tacitly assuming, the magnetic field varies only from right to left, not from front to back, the forces on *BC* and *DA* will be equal and opposite, canceling each other, whereas the force on *AB* will be greater than on *CD* since the field at *AB* is greater. Consequently, the resultant force is from left to right if the induced current is in the direction *ABCD*. This force opposes the original velocity v of the circuit and, therefore, fulfills Lenz's law. Reversing our argument, we can say that the force must be from left to right since it is to oppose the motion and, therefore, the induced current must be in the direction *ABCD* as we have already shown from two other points of view.

16. *E.M.F. Induced by a Moving Magnet.* In the example of induced current we have just been considering, the circuit was moving in a magnetic field that was not uniform. To make this assumption more specific, we can think of the magnetic field as produced by a horseshoe magnet and the motion of the circuit as toward the poles as shown in Fig. 204. It seems fairly obvious that the effects will be the same whether the magnet or the circuit is moved, and this is found to be true experimentally. If the magnetic flux through the circuit is changed, there is an induced e.m.f. no matter what the cause of the change.

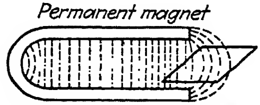


FIG. 204.—A rectangular circuit moving in the nonuniform field of a horseshoe magnet. Changing flux induces an e.m.f.

17. *E.M.F. Induced by a Varying Current.* It is evident that the magnetic flux whose change produces an induced current may come equally well from permanent magnetic poles or from wire-borne currents. In the examples we have been discussing, the strengths of the poles or of the currents setting up the magnetic field have been assumed constant and the change in flux has been the result of motion. In such cases, the induced electric fields could be interpreted in terms of the motion of charges relative to a magnetic field or in terms of changing magnetic flux. But if

the magnetic flux is set up by wire-borne currents, then it can easily be changed by changing the currents without any motion of the conductors involved. In this case, the interpretation in terms of the "motor law" is somewhat forced but in terms of the change in flux is perfectly clear. It is this case that is of the greatest practical importance. Before discussing an important particular example, we shall discuss the effect of using a circuit of several turns for studying the induced current.

E.M.F. Induced in a Coil of N Turns

18. Heretofore, we have been discussing the e.m.f. induced in a rectangular plane circuit of a single wire. This had advantages so long as we were interpreting the induced electric field strength in terms of the motor force on individual electrons, but the shape of the circuit ceases to be of importance as soon as we confine ourselves to the resultant induced e.m.f. caused by the changing flux through the whole circuit. In practice, induced currents are most frequently carried in closed coils of either circular or rectangular cross section. Except in rare instances, these coils have more than one turn. Since the flux passes through each turn, there will be an induced e.m.f. of $d\Phi/dt$ in each turn. If the wire is wound in a continuous spiral always in the same direction, the successive turns are like batteries connected in series so that the resultant e.m.f. is simply the sum of the e.m.fs. of the separate turns. Thus for a coil of N turns, the induced e.m.f. is

$$V = N\frac{d\Phi}{dt}. \tag{5}$$

Air-core Transformer

19. Suppose we have two solenoidal coils A and B set up near each other but forming two separate circuits as in Fig. 205. Any current that flows in coil A will set up a magnetic field. This field will have a definite value at every point in A, depending on the number of windings, the shape of the coil, and the current. The lines of magnetic field strength will be approximately as indicated in the drawing. The magnetic flux through the coil may be called Φ_A, and it will be proportional to the current flowing through A, *i.e.*, $\Phi_A = \mu k_A I_A$ where k_A is a constant of proportion. The constant k_A is the number of turns of the coil per meter if the

coil is sufficiently long and closely wound. The current through A will produce a magnetic field in B so that there will be a magnetic flux through B, resulting from the current in A. This flux Φ_B will, in general, be considerably less than Φ_A since not all the lines of magnetic field produced by A will thread through B, but Φ_B will be proportional to Φ_A and can be written $\Phi_B = C_{AB}\Phi_A$ where C_{AB} is a constant of proportion depending on the relative positions of A and B and never having a value greater than one. In an arrangement of two coils like this, the coil through which the current is originally flowing is called the primary coil and the coil in which the e.m.f. is induced is called the secondary coil. Similarly, the inducing current is called the primary current and the induced current the secondary current.

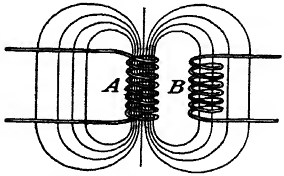


Fig. 205.—A loosely coupled air-core transformer. Changes in the current through A change the flux through B and therefore induce an e.m.f. in B.

20. The two coils just described are said to be coupled. This means that a variation in the current in the primary coil A induces an e.m.f. in the secondary coil B. If the factor C_{AB} is small, the coils are said to be loosely coupled, if it is large, they are said to be closely coupled. The closest coupling possible is obtained by winding the two coils on the same cylinder. In such cases, C_{AB} is very nearly equal to one. If the current in the primary is changing continually, then there will always be an induced e.m.f. in the secondary. This is the situation in a transformer. We shall discuss transformers more fully later on when we have dealt with alternating currents. Except for high frequencies such as are used in radio work, the air-core transformer just described is far less efficient than one with an iron core.

Coefficient of Mutual Inductance

21. We have seen that the flux in the secondary coil is proportional to the flux in the primary coil and that the latter is proportional to the current in the primary coil. The flux in the secondary coil is, therefore, proportional to the current in the primary coil. The induced e.m.f. in the secondary will be given by N times the rate of change of the flux where N is the number of turns and will therefore be proportional to the rate of change

of the current in the primary. Using the calculus notation in which d/dt denotes rate of change with time, we have*

$$E_B = \frac{Nd\Phi_B}{dt} \propto \frac{d\Phi_A}{dt} \propto \frac{dI_A}{dt}$$

or

$$E_B = M_{A} \quad dI_A \tag{6}$$

where M_{AB} is a constant of proportion called the coefficient of mutual inductance. If the rate of change in current is given in amperes per second and the induced e.m.f. in volts, M_{AB} is said to be in henries. Or if we put it another way,

Two coils are said to have a mutual inductance of one henry if a rate of change of current of one ampere per second in one coil produces an induced electromotive force of one volt in the other coil.

This unit like the unit of self-inductance, which we shall discuss presently, is named after Joseph Henry in honor of his pioneer work in this field. Henry was professor of physics at Princeton from 1832 to 1846 and did much of his work on induced currents while there.

Self-inductance

22. We have seen that a variation of the magnetic flux through a coil induces an e.m.f. in that coil. So far we have been assuming that this varying magnetic flux is produced by some outside agency, but it may equally well be produced by a current flowing in the coil itself. For example, suppose that initially there is no current flowing in a coil. There is consequently no magnetic flux. Suppose a switch is closed so that current starts to flow. Immediately, a magnetic field begins to build up in the coil. Any given turn of the coil does not know whether the magnetic field in which it finds itself is produced by other turns of the same coil or by some outside agency. Nor does it care. It immediately tries to resist the building up of the magnetic field in accordance with Lenz's law. This happens in every turn of the coil with the result that there is an induced e.m.f. set up which opposes the

* Earlier in this chapter we used V for potential difference or e.m.f. to avoid confusion with electric field strength. From now on we shall use the more customary E.

increase of current. It takes longer for the current to establish itself than if there were no such effect. Similarly, if there is a steady current flowing in the coil and the circuit is suddenly broken, an e.m.f. is induced in a direction to maintain the original flow of current and original magnetic flux. Such effects are said to result from the "self-inductance" of the coil and are somewhat analogous to inertial effects in mechanics. They are electrical reactions against changes in electromagnetic conditions. The amount of the induced e.m.f. in such a reaction depends on the rate of change of the inducing current and also on the nature of the coil, *i.e.*, on the degree to which the magnetic flux from each turn of the coil goes through all the other turns. Expressed mathematically, we have the induced e.m.f., E, given by

$$E \propto \frac{d\Phi}{dt} = L\frac{dI}{dt} \tag{7}$$

where L is a constant of proportion characteristic of the coil. It is called the coefficient of self-inductance, or merely the self-inductance of the coil, and is a geometrical characteristic of the coil which can be calculated in a few simple cases but is usually determined experimentally.

A coil is said to have a self-inductance of one henry if a rate of change of current of one ampere per second in it produces an induced electromotive force of one volt.

Diamagnetism

23. In our discussion of the air-cored transformer, we mentioned the fact that an iron-cored transformer was far more efficient for low frequencies. We are now ready to consider the whole problem of the effect of materials on induction phenomena. To be specific, suppose we are dealing with primary and secondary coils so closely coupled by winding on the same core that the coupling constant may be considered equal to one. Heretofore, we have considered such coils wound on an air core which has so little effect that it can be neglected. Suppose now that the coils are wound on a solid core of some pure substance. As we have seen in Chap. XV, there exist in such a solid core countless little molecular circuits in which the current flows continually against no appreciable resistance. We saw that the peculiar magnetic properties of iron were explicable in terms of the orientation of

these molecular circuits. But there is another magnetic property of materials that results from the induction of currents in these circuits. Consider in detail the effect of these circuits on the e.m.f. induced in the secondary of our transformer. When the switch in the primary circuit is first closed, the rising current in the primary will induce e.m.fs. in the molecular circuits which, according to Lenz's law, will be in such a direction that the magnetic field of the induced molecular currents will oppose that from the primary current. Since the resistance to the flow of current in the molecular circuits is negligible, the induced molecular currents will continue to flow even after the inducing e.m.fs. have died out so that even after the current in the primary has reached a constant value the molecular currents will be giving a magnetic field opposed to the magnetic field of the primary. If the current in the primary is reduced to zero, the induced e.m.fs. in the molecular circuits will just reduce the currents in them to their original values. Thus the resultant magnetic field for a given current in the primary will be always less than if there were no material present. This means that the resultant change of magnetic flux is less for a given rate of change of current. Consequently, the e.m.f. induced in the secondary by a given rate of change current in the primary is reduced by the presence of the material core.

24. This effect we have been describing, the reduction of magnetic flux by induced molecular currents, is called diamagnetism. It presumably occurs in all materials although in most substances it is masked by the presence of paramagnetic and ferromagnetic effects.

Paramagnetism and Ferromagnetism

25. The diamagnetic effects we have been describing are independent of the orientation of the molecular circuits. As we have seen in Chap. XV, there is another effect that depends on the possibility of orienting the molecular currents. Recalling what was said in Chap. XV, we see that the alignment of these molecular currents depends on the current in the primary coil, not on its rate of change, and that the alignment is in such a direction as to increase the magnetic flux through a coil. In other words, putting iron inside a coil is equivalent to increasing the current through the coil. This effect is called paramagnetism, and when

it exists at all always outweighs the diamagnetic effect. In most paramagnetic substances, the effect is not very large, usually increasing the magnetic flux by a small percentage. In some substances, notably iron, nickel, and cobalt, the effect is so much larger as to warrant a different name, ferromagnetism. Such substances are called ferromagnetic substances.

Relative Permeability

26. We have described the effect of molecular circuits in increasing or decreasing the resultant magnetic flux from wire-borne currents. It is convenient to be able to express this effect quantitatively and in a way that is independent of our theoretical interpretation of the observed phenomena. To do this, we can compare the e.m.fs. induced in the secondary of a closely coupled transformer when cores of various substances are used. These results can then be referred to some standard. Strictly speaking, the proper standard is the e.m.f. induced when the coils are in a vacuum, but the effect of air is so small that it can be neglected for all but the most accurate measurements, and therefore we shall take an air core as our standard. According to Par. 21 on mutual inductance, the e.m.f. in the secondary is

$$E = M\frac{dI}{dt}$$

where M is the mutual inductance of the two coils. M depends on the change of flux produced in the secondary by a given change of current in the primary and, therefore, depends on both the geometry of the coils and the paramagnetic and diamagnetic effects of material in the neighborhood. Let M_0 be the value of the mutual inductance in the absence of any magnetic material and M be the value when there is magnetic material present. Then the ratio of these two quantities is a measure of the effect of the magnetic material. In general, the magnetic field of the primary will extend throughout space, and therefore magnetic material anywhere in space will affect the flux through the secondary produced by a given current in the primary. It is obviously inconvenient to fill the whole of space with the material to be tested. We could choose any arbitrary arrangement of coils and material as a standard way of measuring and expressing the effect of a magnetic material. But there is one particular arrangement

that is peculiarly suitable. It can be shown both experimentally and theoretically that the magnetic field of a ring-shaped solenoid such as shown in Fig. 206 is zero everywhere outside the solenoid (see Chap. XVI, Par. 9). Therefore, if such a coil is used as a primary with a secondary wound around all or part of it, the full magnetic effect of any material may be obtained by filling the space inside the coils with an annular ring of the material. The mutual inductance of such an arrangement can then be compared with that of identical coils with air cores. If M_0 is the value of the mutual inductance for an air core and M for the solid core, M/M_0 is called the relative permeability of the substance and is denoted by μ_r. It is clear that μ_r is also the ratio of the magnetic flux in the substance to the flux in air. For diamagnetic sub-

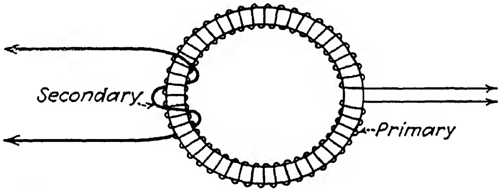


FIG. 206.—A ring solenoid with a primary winding of many turns and a secondary winding of a few turns.

stances, μ_r will be slightly less than one; for paramagnetic substances, somewhat greater than one; and for ferromagnetic substances, very much greater than one, perhaps as much as two or three thousand.

27. Evidently the reason for the permeability of materials differing from that of empty space and differing among each other is to be found in the effect of the submicroscopic currents such as were discussed in Chap. XV when we first introduced the subject of magnetism and again in Par. 25 above. Thus in a solenoid with an iron core, the effect of all these atomic currents is added to that of the current in the solenoid to give a resultant magnetic flux much greater than without the iron, perhaps thousands of times greater. There are two ways in which we can handle this situation. In the first, we simply think of the iron-cored solenoid as equivalent to an air-cored solenoid with a greatly increased number of ampere turns per meter. In other words, we interpret the observations in terms of atomic structure and conclude that

the alignment of the atomic circuits produces current additional to that in the winding of the solenoid so that the total effective current is multiplied enormously. This is the most satisfactory way of considering the situation from the point of view of understanding. But from the practical point of view, a different way of looking at the thing is more useful. We think of the magnetic field strength as constant, given by the number of ampere turns per meter of the solenoid windings, but we say that the magnetic induction or flux that can be produced by this field depends on the nature of the material in the field. This dependence is taken care of by the permeability μ whose name itself carries the historical idea that certain substances were more permeable to magnetism than others. This second method of dealing with the effects of magnetic materials stresses the resultant magnetic induction rather than the magnetic field causing the induction. Correspondingly, we shall find it somewhat more convenient for the rest of this chapter to use B instead of μH in our equations.

Magnetic Saturation

28. It is evident that the greatest possible contribution that the molecular currents in iron can make to the flux through a circuit is made when they are all aligned with their axes parallel to the field. As this condition is approached, the aligning effect of increasing the current in the coil becomes less and less marked and the ratio of flux to wire-borne current decreases. The permeability, therefore, is not a constant quantity but falls off as the iron becomes more highly magnetized. Finally, a condition of maximum flux is reached where further increase in the current in the primary coil has no effect. This is called a condition of magnetic saturation.

Hysteresis

29. If the magnetic field is decreased after saturation, the alignment of the molecular circuits will tend to persist, remaining appreciable even when the magnetic field is reduced to zero. Consequently, the B vs. H curve will be quite different with rising and with falling H. Furthermore, if H falls to zero and then increases in the negative direction, it must reach an appreciable value to break down the tendency of the molecular circuits to keep themselves aligned in the positive direction. Thus B

will remain positive until H has a fairly large negative value. The same sort of lag will appear as H is made positive again. This process is called hysteresis, and curves like that in Fig. 207 are called hysteresis curves. The exact shape of such a curve depends on the exact composition and heat treatment of

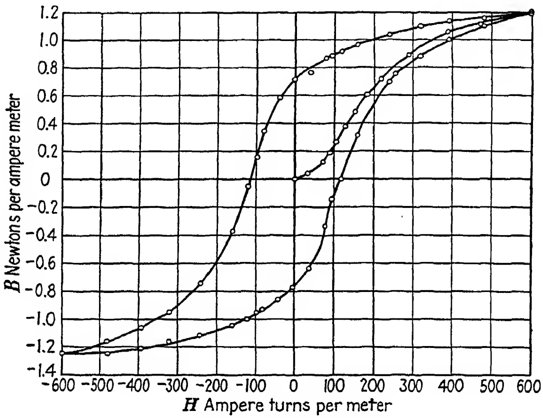


FIG. 207.—A magnetization curve and hysteresis loop showing the relation of B to H under varying conditions. H is increased from zero in the positive direction giving the branch of the curve starting from the origin. It is then decreased, reversed, increased negatively, etc. and the corresponding values of B observed.

the magnetic material and is of great importance in transformer design.

Rotating Coils

30. In introducing the idea of induced currents, we confined our examples of circuits moving in magnetic fields to translational motion. In practice, a more important case is that of the rotation of a circuit in a magnetic field. Suppose we consider a simple rectangular circuit mounted so that it can turn around an axis perpendicular to a uniform magnetic field. Such a circuit is shown in Fig. 208. The magnetic flux through the circuit varies from a maximum when the plane of the circuit is perpendicular to the direction of the magnetic field to zero when the plane of the circuit is parallel to the magnetic field. At any intermediate position, the flux will be given by

$$\Phi = BA \sin \theta \tag{8}$$

where B is the magnetic induction, A the area of the circuit, and θ the angle between the direction of the field and the plane of the circuit. If the circuit is rotating with an angular velocity ω, then the induced e.m.f. at any instant is, by differentiation,

$$\frac{d\Phi}{dt} = BA(\cos \theta)\frac{d\theta}{dt} = BA\omega \cos \theta. \tag{9}$$

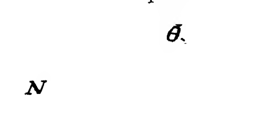


FIG. 208.—A plane coil rotating about an axis perpendicular to a uniform magnetic field. As the coil turns the flux through it alternates from a maximum in one direction ($\theta = 90°$) through zero ($\theta = 180°$) to a maximum in the other direction ($\theta = 270°$) and back to zero ($\theta = 0°$).

(To obtain this result without the use of the calculus, we can proceed as follows: Suppose that in the time Δt the angle θ changes by $\Delta\theta$, then by definition $\Delta\theta/\Delta t = \omega$, the angular velocity. In this same time interval, the flux changes from $BA \sin \theta$ to $BA \sin (\theta + \Delta\theta)$ so that its rate of change is

$$\frac{BA[\sin (\theta + \Delta\theta) - \sin \theta]}{\Delta t} = \frac{BA(\sin \theta \cos \Delta\theta + \cos \theta \sin \Delta\theta - \sin \theta)}{\Delta t},$$

but if $\Delta\theta$ is very small, $\cos \Delta\theta = 1$ and $\sin \Delta\theta = \Delta\theta$; therefore the preceding equation reduces to

$$\frac{BA(\sin \theta + \Delta\theta \cos \theta - \sin \theta)}{\Delta t} = BA \cos \theta \frac{\Delta\theta}{\Delta t} = BA\omega \cos \theta$$

which is the result just obtained by the use of the calculus.) Assuming $\theta = 0$ when $t = 0$, we can write for θ at any subsequent time t, $\theta = \omega t$. Furthermore, if our rotating coil has N turns, the induced e.m.f. is equal to N times the rate of change of flux. Making these substitutions, we have for the e.m.f. of a coil of N turns and area A, rotating in a magnetic field H with an angular velocity ω,

$$E = NBA\omega \cos \omega t = E_0 \cos \omega t \tag{10}$$

where

$$E_0 = NBA\omega \quad \text{or} \quad N\mu HA\omega. \tag{11}$$

If we graph the induced e.m.f. against the time, we get a curve such as is shown in Fig. 209. Evidently the induced e.m.f. not only changes in magnitude but alternates in direction. This alternation not only is shown by the equation but can be confirmed by applying our simple rules. According to the equation, the induced e.m.f. changes sign when $\omega t = \theta = 90°$, but this is the point where we can see from Fig. 209 that the flux through the circuit ceases to increase and starts to decrease so that the induced e.m.f. changes from one opposing the flux to one reinforcing it.

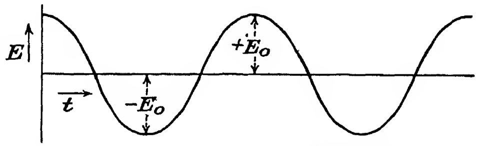


FIG. 209.—The e.m.f. induced in the coil of Fig. 208 by the alternating flux. It is assumed that $\theta = 0$ when $t = 0$.

Such an e.m.f. as shown in Fig. 209 is called an alternating e.m.f., and the current corresponding to it is called an alternating current.

Direct- and Alternating-current Generators

31. The e.m.fs. generated in a coil like that just described can be used to drive currents through external circuits. Connections can be made in two different ways. The simplest way is to have the two ends of the coil connected to rings on the shaft that are connected to the external circuits through sliding contacts. The e.m.f. in the external circuit is then like that in Fig. 209, and we have an alternating-current, or a-c, generator. Or the ends of the coil can be connected to the two halves of a single split ring on the shaft as shown in Fig. 210. This is called a commutator and acts like an automatic reversing switch so that just as the induced e.m.f. in the rotating coil reverses, the brushes C and D lose contact with A and B, respectively, and make contact with B and A. The e.m.f. in the external circuit is, therefore, always in the same direction but varies in magnitude as shown in Fig. 211. If instead of a single rotating coil, a whole group of rotating coils is used whose planes make equal angles with each other, and whose terminals come out to a more complicated commutator, an

e.m.f. can be obtained which is almost constant. Plotted against time the potential between the brushes of a three-coil generator is as shown in Fig. 212. A still further increase in the number of coils makes the e.m.f. more nearly constant, so much so that the

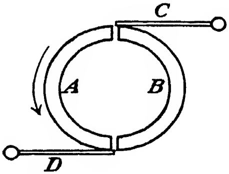


FIG. 210.—A commutator. The terminals of the coil of Fig. 208 are connected to A and B and therefore alternately to C and D.

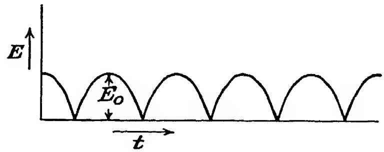


FIG. 211.—E.m.f. in an external circuit connected through a commutator (Fig. 210) to a single rotating coil (Fig. 208). The commutator is set so that the connections to the ends of the coil are reversed just as the coil goes through the position of maximum flux, *i.e.*, where the rate of change of flux changes sign.

current driven by it through an external circuit will not vary appreciably. Such a current is called a direct current and such a generator a direct-current, or d-c, generator.

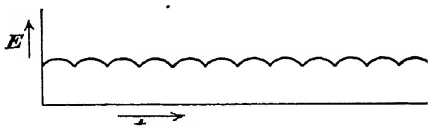


FIG. 212.—Approximately constant d.-c. e.m.f. from a generator having three plane coils at angles of 60° connected to a commutator of six segments.

SUMMARY

According to the motor law, electrons in a wire that is moving perpendicular to a magnetic field are subjected to a force along the wire equivalent to that caused by a potential difference between the ends of the wire. The resultant of such equivalent potential differences added up around a closed circuit is called the induced e.m.f. in the circuit. If the magnetic flux is defined as the product of the area of the circuit and the component of magnetic induction perpendicular to it, the induced e.m.f. is always proportional to the rate of change of magnetic flux whether or not that change results from motion of parts of the circuit. Expressed mathematically, $E = d\Phi/dt$ where E is the induced

e.m.f. and Φ is the magnetic flux. According to Lenz's law, the induced current in a circuit will always flow in a direction to retard the change of flux. For a coil of N turns the induced e.m.f. is $Nd\Phi/dt$.

A transformer consists of two coils in separate circuits but close enough together so that a current in one sets up a considerable magnetic flux in the other. The coils being called A and B, a varying current I_A in A evidently causes a varying flux and, therefore, an induced e.m.f. in B. This e.m.f. E_B is proportional to the rate of change of I_A, *i.e.*, $E_B = M_{AB}dI_A/dt$, where the factor of proportion M_{AB} is called the coefficient of mutual inductance. Changing current in a coil induces an e.m.f. in that coil itself as well as in neighboring coils. Such an e.m.f. is given by $E = LdI/dt$, where L is called the coefficient of self-inductance of the coil.

The resultant magnetic flux set up by a current in a coil depends on the material of the core around which the coil is wound. "Diamagnetic" materials reduce the flux slightly, "paramagnetic" materials increase it slightly, and "ferromagnetic" materials increase it to as much as several thousand times its value when the coil has no core. Diamagnetism is interpreted as resulting from induced currents in molecular circuits, paramagnetism and ferromagnetism as resulting from the orientation of molecular circuits. These effects are given quantitatively by the relative permeability of a material.

In ferromagnetic substances, the permeability decreases with increasing magnetic field as the molecular circuits approach complete alignment. Furthermore, the permeability for increasing field strength is less than for decreasing field strength of the same value. This is shown by B vs. H curves known as hysteresis curves.

A plane coil rotating in a constant magnetic field generates an alternating e.m.f. in its windings. This is the principle of all d-c and a-c generators.

ILLUSTRATIVE PROBLEMS

1. A coil of one turn 30 cm. in diameter is lying flat on the floor at a place where the vertical component of the earth's magnetic field is 50 amp. turns/m. If the coil is turned over in $\frac{1}{4}$ sec., what average e.m.f. will be induced in it?

The area of the coil is $\pi d^2/4 = \pi(0.3 \text{ m.})^2/4 = 0.0706 \text{ m.}^2$ The total flux

$$\Phi = BA = \mu_0 HA = 4\pi \times 10^{-7}\frac{\text{newton}}{\text{amp.}^2} \times 50\frac{\text{amp. turns}}{\text{m.}} \times 0.0706 \text{ m.}^2$$
$$= 4.44 \times 10^{-6} \text{ joule/amp.}$$

The average induced e.m.f. is equal to the rate at which the flux through the coil changes. When the coil is being turned over, the flux through it, due to the vertical component of the earth's magnetic field, drops to zero after a rotation of 90°. The flux then builds up to its original value, but in the opposite direction through the coil, in the next 90°. Therefore, the total change in the flux through the coil in turning it over is twice the original value of the flux. The average induced e.m.f. is, therefore, Eq. (4), page 448,

$$V = \frac{\Delta\Phi}{\Delta t} = \frac{2 \times 4.44 \times 10^{-6} \text{ joule/amp.}}{\frac{1}{4} \text{ sec.}}$$

$$V = 3.55 \times 10^{-5}\frac{\text{volt-coulomb}}{\text{coulomb}} = 3.55 \times 10^{-5} \text{ volt.}$$

2. Two coils are placed side by side so that they have a mutual inductance of 5 millihenries. If a current of 10 amp. is reversed in 0.01 sec. in one coil, what is the average induced e.m.f. in the other coil?

Equation (6), page 454, gives the e.m.f. induced in one coil due to the rate of change of current in another coil. Thus,

$$E_B = M_{AB}\frac{dI_A}{dt} \tag{12}$$

where E_B = the e.m.f. induced in coil B due to the rate of change of current in coil A.

$M_{AB} = 5 \times 10^{-3}$ henry = the mutual inductance of the two coils.

dI_A = 20 amp. = the total change in the current in the coil A.

dt = 0.01 sec. = the time during which the current changes.

Substituting these values in Eq. (12) gives

$$E_B = 5 \times 10^{-3} \text{ henry} \times \frac{20 \text{ amp.}}{0.01 \text{ sec.}} = 10\frac{\text{volt-sec.}}{\text{amp.}} \times \frac{\text{amp.}}{\text{sec.}} = 10 \text{ volts.}$$

PROBLEMS

1. A rectangular circuit with the side AB free to slide along parallel to itself is placed in a magnetic field of 5×10^5 amp. turns/m. perpendicular to its plane. If AB is 10 cm. long and moves with a uniform velocity $v = 3$ m./sec., what is the induced e.m.f. in the circuit in volts? If the circuit has a resistance of 3 ohms, how much current will flow? If the magnetic field is into the paper, what is the direction of the current?

2. If, in Prob. 1, the plane of the circuit is vertical, with the magnetic field still perpendicular to the plane of the circuit, and AB falls freely

under gravity, what will be the average e.m.f. induced in the circuit during the fall? *AB* starts at rest with *AD* equal to 25 cm. (Fig. 213.)

3. A propeller in an airplane is 2m. long and makes 4,000 r.p.m. If the propeller is covered with metal and rotates in a plane perpendicular to the earth's magnetic field ($H = 50$ amp. turns/m.), find the difference of potential between the axle and the tip of the propeller blade.

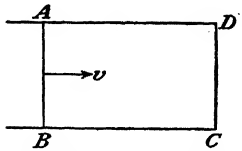


FIG. 213.—Rectangular loop of wire with a movable side. Probs. 1 and 2.

4. A rectangular loop of wire 40×60 cm. is rotated with constant angular velocity $\omega = 40$ rad./sec. about an axis parallel to its long sides and half way between them. The loop is in a uniform magnetic field of 2×10^5 amp. turns/m. with its axis perpendicular to the field. Obtain the e.m.f. as a function of the time. If connections are made to the loop through slip rings to an external circuit whose resistance is 50 ohms, what is the current in the latter when the lines of force lie in the plane of the loop? (Neglect the resistance of the loop.) What is the current when the plane of the loop makes an angle of 45° with the lines of force?

5. If two coils are coupled so that their mutual inductance is 0.02 henry, what is the e.m.f. induced in one at the instant that the current is changing at a rate of 30 amp./sec. in the other?

6. A coil has a self-inductance of 100 millihenries and a resistance of 3 ohms. If the e.m.f. impressed across the coil instantaneously is 90 volts and the current (in the same direction) is increasing at a rate of 75 amp./sec., what is the instantaneous value of the current through the coil?

7. A bar magnet fits closely inside a flat coil of wire of 150 turns. If the total flux emanating from the N pole of the magnet is 80 webers and the magnet is pulled completely out of and away from the coil in 0.1 sec., what is the average e.m.f. induced in the coil?

8. A coil of 50 turns with a radius of 3 cm. is placed between the poles of an electromagnet with its plane normal to the lines of force. The coil is connected in series with a circuit of resistance 480 ohms, its own resistance being 20 ohms. When the coil is jerked suddenly out of the field in 0.05 sec. it is found that the average induced current through it is 0.30 milliamp. Compute the field strength of the electromagnet.

9. A circuit consists of a circular coil of 15 turns of flexible wire 30 cm. in diameter lying in a horizontal plane. This circuit is pulled out straight so that the two sides touch each other in $\frac{1}{10}$ sec. at a place where the vertical component of the earth's magnetic field is 50 amp. turns/m. What is the average e.m.f. induced in the coil?

10. A vertical magnetic field varies at the rate of 100 amp. turns/m./cm. in a given direction. A horizontal coil of 50 turns 300 sq. cm. in area moves with a constant velocity of 5 m./sec. in the given direction. What e.m.f. will be generated in the coil?

11. A coil of N turns of area A sq. m. falls freely under gravity with its plane vertical. If it is in a region where a horizontal magnetic field is increasing in a direction vertically downward at the rate of H amp.

turns/m./m., what will be the instantaneous e.m.f. induced in the coil after it has fallen a height L m. from rest?

12. At what rate is the instantaneous induced e.m.f. in the coil of Prob. 11 increasing?

13. A coil of 20 turns of area 100 sq. cm. has a maximum e.m.f. of 1 volt induced in it by rotating around an axis perpendicular to a magnetic field of 10^3 amp. turns/m. What is the angular velocity of the coil?

14. Two coils of 20 turns and 400 sq. cm. area are mounted with their planes perpendicular to each other on an axle along the line of intersection of their planes. The axle is perpendicular to a uniform magnetic field of 10^4 amp. turns/m. and rotates with an angular velocity of 10^4 r.p.s. The two coils are connected in series. Plot the e.m.f. induced in the two coils as a function of the time, starting with one coil perpendicular to the magnetic field.

15. Two coils are wound in a ring shape as in Fig. 206. With an air core, an e.m.f. of 10^{-4} volt is induced in the secondary when the current in the primary changes at the rate of 0.1 amp./sec. With an iron core, an e.m.f. of 1 volt is induced in the secondary when the current in the primary changes at the rate of 0.5 amp./sec. What is the permeability of the iron?

CHAPTER XXIII

MECHANICAL AND ELECTRICAL OSCILLATIONS

1. In studying mechanics in the early chapters of this book, we confined our attention to bodies moving with constant accelerations in response to the action of constant forces. Many problems of mechanics and electricity fall into this simple category. But the problem of a charged particle moving in a magnetic field was of a different type involving the kind of force that causes motion in a circle. We therefore studied such motion and the kind of force producing it, a force constant in magnitude but varying in direction according to a simple law. Obviously, there must be many cases, in nature, of motion under the action of forces that are varying in both direction and magnitude. We cannot deal with any very general case, but we can consider the motion caused by the action of a force that varies according to one particular very simple law; and we shall find that this case has many important applications, both mechanical and electrical.

Definition of Simple Harmonic Motion

2. In Chap. IV, Par. 20, we discussed the conditions for the equilibrium of a body, the conditions under which a body would remain motionless. We did not differentiate between different kinds of equilibrium, a subject probably already familiar. Reviewing it briefly, we may recall that a body is said to be in stable equilibrium if it tends to return to its original position after it has been slightly displaced, in unstable equilibrium if the displacement tends to increase, and in neutral equilibrium if the displacement does not affect the equilibrium. The three possible positions of a cone are the classical examples of the three states of equilibrium. Standing on its base the cone is stable; balanced on its point it is unstable; and lying on its side it is in neutral equilibrium.

3. We are concerned with the particular case of stable equilibrium where a small displacement not only introduces a force

tending to bring the body back to its original position but where that force is exactly proportional to the amount of the displacement. Suppose, for example, that we have a weight suspended by a spring. Initially, it is at rest. We pull the weight down against the elastic restoring force of the spring. We know that the force resisting this distortion of the spring is proportional to the distortion and in the opposite direction to it (Hooke's law). If we call the displacement of the weight from its equilibrium position x, then the restoring force will always be $-kx$ where k

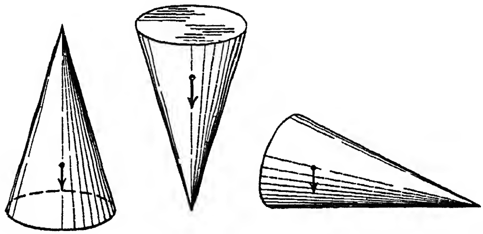


FIG. 214.—Three positions of a cone illustrating stable, unstable and neutral equilibrium.

is a factor of proportion characteristic of the spring. If the weight is suddenly released when it is in a displaced position, the restoring force will accelerate it toward its position of equilibrium and it will oscillate about this position until the frictional forces finally reduce its motion to zero. Motion of this type is called simple harmonic motion and may be precisely defined as follows:

Simple harmonic motion is the oscillatory motion of a body about a position of equilibrium toward which the body is drawn by a restoring force proportional to its displacement from that position.

In mathematical language, if x is the coordinate defining the position of the body ($x = 0$ when the body is at rest) and F is the force on it, the condition for simple harmonic motion is that $F = -kx$, where k is a constant of proportion.

4. We now wish to derive equations that will tell us about simple harmonic motion as completely as the equations in Chap. XVIII, Par. 11, tell us about linear motion and circular motion. By far the simplest and most elegant way of doing this is by the use of the calculus. It is possible, however, to avoid the calculus, and a second treatment without its use will also be given.

Calculus Treatment of Simple Harmonic Motion

5. We know, in general, that the acceleration is the rate of change of velocity and that velocity is the rate of change of position or, in the notation of calculus, $a = dv/dt$ and $v = dx/dt$. From this, it is clear that

$$a = \frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}$$

Therefore our condition for simple harmonic motion reduces to

$$F = ma = m\frac{d^2x}{dt^2} = -kx \qquad \text{or} \qquad \frac{d^2x}{dt} = -\frac{k}{m}x. \tag{1}$$

Our problem might now be considered a purely mathematical one, namely, to find a relation between x and t such that the second derivative of x with respect to t will be proportional to x itself. If we can do this, we can tell what value x has for any given value of t and also what value v and a have for any given value of t. In other words, we can describe the motion completely. Equation (1) is a simple example of what is known as a differential equation and can be solved by perfectly general mathematical methods. Unfortunately this would overtax the mathematical training of most beginning students. We must proceed by a less clean-cut method depending on physical reasoning.

6. Consider the example of simple harmonic motion cited in Par. 3, a weight oscillating up and down on a spring. It is evident that x, the displacement from the position of equilibrium, has both positive and negative values and oscillates between approximately equal positive and negative extremes as time goes on. No matter how large t becomes, x never exceeds a certain value and every value of x repeats itself periodically. Suppose we try to express this simple observation by a mathematical relation between x and t. There are two familiar relations between variables that have this property, namely, sine and cosine of an angle; each of these functions varies regularly between plus and minus one as the angle is increased indefinitely. Consider the equation

$$x = c \cos \omega t \tag{2}$$

where c and ω are constants as yet undetermined. Evidently it

satisfies our physical requirements, as far as position goes; it gives a variation of x between c and $-c$ as t increases indefinitely and every value of x repeats itself every time t increases by $2\pi/\omega$. Let us now see whether it also satisfies our law of force as expressed mathematically by Eq. (1). Differentiating (2), we get

$$\frac{dx}{dt} = -c\omega \sin \omega t. \tag{3}$$

Differentiating again and substituting from (2), we get

$$\frac{d^2x}{dt^2} = -\omega^2 c \cos \omega t = -\omega^2 x. \tag{4}$$

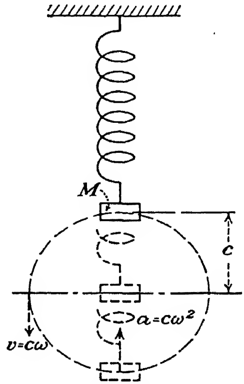


FIG. 215.—Simple harmonic motion of a weight on the end of a spring. At the top position $x = c$, $v = 0$, $a = -c\omega^2$; at the middle position $x = 0$, $v = \pm c\omega$, $a = 0$ and at the bottom position $x = -c$, $v = 0$ and $a = c\omega^2$.

Evidently Eq. (2) relating x and t is consistent with Eq. (4) between d^2x/dt^2 and x; in other words, Eq. (2) is a "solution" of the differential Eq. (4). But Eq. (4) is identical with Eq. (1) if k/m be substituted for ω^2. Consequently, if we substitute $\sqrt{k/m}$ for ω in (2), we have the solution of (1) for which we have been seeking. Making this substitution in (2), (3), and (4) and collecting our equations, we have

$$\text{Force} \qquad F = -kx \text{ (by definition)} \tag{5a}$$

$$\text{Displacement } x = c \cos \sqrt{\frac{k}{m}}\,t \tag{5b}$$

$$\text{Velocity} \qquad v = -c\sqrt{\frac{k}{m}} \sin \sqrt{\frac{k}{m}}\,t \tag{5c}$$

$$\text{Acceleration} \quad a = -c\frac{k}{m} \cos \sqrt{\frac{k}{m}}\,t \tag{5d}$$

$$\text{Period} \qquad T = 2\pi\sqrt{\frac{m}{k}} \text{ (see Par. 8).} \tag{5e}$$

These equations give not only the magnitudes of F, x, v, and a but also their direction. They are based on the convention that the original displacement is positive and positive values of F, v, and a are heading in this direction, negative in the opposite direction.

Phase, Amplitude, Period, and Frequency

7. It is of interest to rewrite these equations in simplified form setting $\theta = \omega t = \sqrt{\frac{k}{m}}t$ and remembering that $-\sin\,\theta = \cos\left(\theta + \frac{\pi}{2}\right)$ and $-\cos\,\theta = \cos\,(\theta + \pi)$

$$x = c\cos\omega t \qquad x = c\cos\theta \tag{6a}$$

$$v = c\omega\cos\left(\omega t + \frac{\pi}{2}\right) = c\omega\cos\left(\theta + \right. \tag{6b}$$

$$a = c\omega^2\cos\,(\omega t + \pi) = c\omega^2\cos\,(\theta + \pi). \tag{6c}$$

Thus we see that all three quantities vary as the cosine of an angle and that the variable part of the angle is the same in each case. In periodically varying quantities of this sort, the angle variable is called the phase. Thus the phase of x at the time t is ωt, the phase of v, $\omega t + \frac{\pi}{2}$, and the phase of a, $\omega t + \pi$. The constant angles added to θ in the case of v and a are called phase angles. x and v are said to differ in phase by $\pi/2$ rad. or 90°; similarly, a and x differ in phase by 180°. The significance of these angles may be clearer after the second treatment of simple harmonic

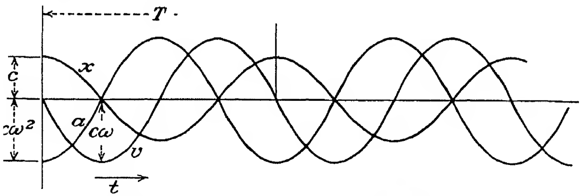


FIG. 216.—The displacement x, velocity v, and acceleration a, plotted against time for a simple harmonic motion of amplitude c and period $T = 2\pi/\omega$. The time is measured from an instant when $x = c$.

motion in Par. 9. The differences in phase of the three quantities can also be seen in the graphs of Fig. 216. Physically, a difference in phase between two periodic quantities means that they reach their maximum values at different times. The one reaching its maximum first is said to lead the other in phase. The second quantity is said to lag behind the other in phase. If the time is measured from some arbitrary zero not corresponding to a maximum value of any of the periodic quantities in a problem, then

the equation for each contains an arbitrary angle. For example, we would have $x = c \cos (\omega t + \phi)$ where ϕ would be called the phase angle or epoch of x. It is seen from the equations or from the foregoing graph that c is the *maximum value that x ever reaches.* It *is called* the *amplitude* of the motion or the amplitude of x. Similarly, the amplitude of v is $c\omega$ and of a is $c\omega^2$.

8. It is often of interest to determine how long it takes for the motion to repeat itself, *i.e.*, how long an interval elapses between the time that x, v, and a have certain values and the time that they have the same values again. Clearly, this is the time required for θ to increase by 2π. Calling this time the period and denoting it by T, we have

$$\omega(t + T) = \theta + 2\pi \qquad \text{where} \qquad \omega t = \theta$$

or

$$\omega T = 2\pi \qquad \text{or} \qquad T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}} \tag{7}$$

We also sometimes are interested in *the number of times a second that the body oscillates, i.e., the reciprocal of the period.* This *is called the frequency* and denoted sometimes by f and sometimes by the Greek letter ν (nu). One of the most interesting characteristics of simple harmonic motion is that the period is independent of the amplitude. This is brought out by Eq. (7) which shows that T does not depend on c. Physically, this means that a body oscillating with simple harmonic motion travels farther if the amplitude increases but it also travels just enough faster on the average to compensate for the increase in distance. This independence of the period and amplitude was first noticed toward the end of the sixteenth century by Galileo watching the swinging chandelier in the cathedral at Pisa.

Simple Harmonic Motion in Terms of Rotation

9. It happens that we can get a simple and accurate description of simple harmonic motion in terms of circular motion. If we consider a point moving around the circumference of a circle with uniform speed, then we find that the projection of this point on a diameter of the circle moves with simple harmonic motion. This is clear from Fig. 217 where the point P is moving with constant angular velocity ω and P' is its projection on a diameter. The

position of P at any time is determined by θ and the position of P' by OP' which we may call x. Clearly, $OP' = c \cos \theta$ if c is the radius of the circle; but the value of θ at any time is given by $\theta = \omega t$ if we start measuring the time when $\theta = 0$. Therefore, we have $x = c \cos \omega t$ which is exactly the form we had in Eq. (2) for the displacement in simple harmonic motion. We can show, furthermore, that the velocity and acceleration of P' also are of the same form as in simple harmonic motion. This can be done most easily by the calculus, as before; but for the benefit of those that have not studied calculus, we shall do it by a method equally exact but more cumbersome. Draw a vector at P representing the linear velocity of P which will be ωc. The velocity of P' will be the horizontal component of this velocity, *i.e.*, $\omega c \sin \theta$. Similarly, the acceleration of P' will be the horizontal component of the acceleration of P; but the latter is $\omega^2 c$ directed in along the radius so that the acceleration of P' is $\omega^2 c \cos \theta$. Both the velocity and the acceleration are in the direction of decreasing x and should, therefore, be written with a negative sign. The motion evidently repeats itself every time P moves completely around the circle, *i.e.*, for every increase of θ by 2π. Therefore, the period is $2\pi/\omega$. We have, then, for the displacement, velocity, acceleration, and period of the point P' the following equations:

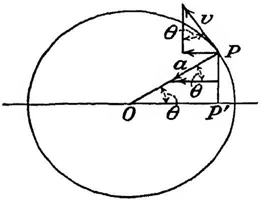


FIG. 217.—Simple harmonic motion in terms of circular motion. The point P moves with uniform angular velocity around the circle. P', the projection of P on a diameter of the circle, executes simple harmonic motion.

$$x = c \cos \omega t \tag{8a}$$

$$v = -c\omega \sin \omega t = c\omega \cos \left(\omega t + \frac{\pi}{2}\right) \tag{8b}$$

$$a = -c\omega^2 \cos \omega t = c\omega^2 \cos (\omega t + \pi) \tag{8c}$$

$$T = \frac{2\pi}{\omega} \tag{8d}$$

which are the same as Eq. (5) except that $\sqrt{k/m}$ has been replaced by ω.

10. Apparently, the vector OP of length c rotating counterclockwise with the angular velocity ω or frequency $f = \omega/2\pi$,

represents a simple harmonic motion of amplitude c and frequency $f = \omega/2\pi$. The displacement, velocity, and acceleration of the corresponding simple harmonic motion can be obtained at any time by using Eqs. (8) which are seen to depend only on the length c of the rotating vector, on its angular velocity, and on the time (clearly c is the amplitude of the motion as before). Evidently the phase of the periodic motion is the angle that OP makes with the axis OP'. In the preceding discussion, the phase angle ϕ has been taken as zero. For some purposes, it is best to use the equations for a simple harmonic motion in one form, for others in another. Several forms have already been given. The student is urged to write the equations out first using the frequency instead of the angular velocity and then using the period instead of the angular velocity.

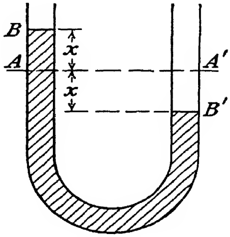


FIG. 218.—Liquid executing simple harmonic motion in a U tube.

Some Examples of Simple Harmonic Motion

11. Suppose that we have some liquid in a U tube of uniform cross section. If undisturbed, the liquid will be in stable equilibrium, standing at the same height in the two sides of the tube. If the liquid is momentarily displaced, by blowing down one side of the tube for instance, it will oscillate around its position of equilibrium. This oscillation is a simple harmonic motion as we can easily see. In Fig. 218, let AA' be the equilibrium position of the liquid. Let B and B' indicate a displaced position of the liquid. Let x be the distance $AB = A'B'$ which gives the amount of displacement from the position of equilibrium. Then the force tending to restore the liquid to its equilibrium position at any time is the weight of the column of liquid of length AB plus $A'B'$ or $2x$. If the density of the liquid is d and the cross section of the tube is S, then the restoring force is $-2Sdgx$. Since S, d, and g are constant, the restoring force is proportional to the displacement. This is the condition for simple harmonic motion, and, therefore, the liquid will oscillate about its position of equilibrium according to the laws of simple harmonic motion deduced above. The equation for the force written in general terms was $F = -kx$ and this now becomes $F = -2Sdgx$. In

other words, the constant of proportion k between the force and the displacement is in this particular case $2Sdg$. Furthermore, the mass of the moving body is simply the mass of the whole length of the liquid. If we let L be this length, *i.e.*, the distance from A around the U to A', the mass of the liquid is LSd. The equations governing the motion of the liquid are then obtained by substitution for k and m in Eq. (5) above. The $\sqrt{k/m}$ in these equations becomes

$$\sqrt{\frac{2Sdg}{LSd}} = \sqrt{\frac{2g}{L}}.$$

and we have

$$\left.\begin{aligned}
&\text{Force} && F = -2Sdgx \\
&\text{Displacement} && x = c\cos\sqrt{\frac{2g}{L}}\,t \\
&\text{Velocity} && v = -c\sqrt{\frac{2g}{L}}\sin\sqrt{\frac{2g}{L}}\,t \\
&\text{Acceleration} && a = -c\frac{2g}{L}\cos\sqrt{\frac{2g}{L}}\,t \\
&\text{Period} && T = 2\pi\sqrt{\frac{L}{2g}}.
\end{aligned}\right\} \tag{9}$$

In these equations, the amplitude c will depend on the strength of the original disturbance which sets the liquid in motion. If this disturbance is a simple displacement, c is the magnitude of this initial displacement. As is apparent from the preceding equations, the period is independent of the amplitude. In discussing this example as in the case of the earlier general discussion, we have neglected the effect of friction.

12. *Weight on a Spiral Spring.* In defining simple harmonic motion, we pointed out that any elastic system that satisfied Hooke's law would execute simple harmonic motion, and we spoke specifically of a weight suspended by a spiral spring. It is not worth our while to express the constant k in terms of the shape of the spring and the elastic constants of the material of which it is made, and therefore Eqs. (5) already are sufficiently explicit for this case if we interpret k as expressing the rigidity of the spring and m as the mass of the weight hung from it (corrected if necessary for the mass of the spring). It is evident that the period becomes shorter the stiffer the spring and longer the

greater the weight. These conclusions can be easily verified by experiment.

13. *Torsion Pendulum.* This case of simple harmonic motion involves so many questions that we have barely touched upon that we shall treat it in detail. First, let us specify exactly what we mean by a torsion pendulum. We mean a weight of some sort hung at the end of a straight vertical wire and fastened rigidly to the wire so that any rotation of the weight twists the wire (Fig. 219). This twist resists the rotation and there is a position of equilibrium where there is no twist of the wire. A rotation in either direction from this position is resisted by the elastic torque of the wire. Furthermore, this elastic torque obeys Hooke's law, increasing in magnitude as the angular displacement from the equilibrium position increases. We have already encountered this sort of an arrangement in Cavendish's determination of the gravitational attraction and in Coulomb's experiment on the law of force between electric charges (Chap. IV, Par. 10, and Chap. X, Par. 10). In each of these cases, the suspended weight was a long crossbar with metallic spheres at its ends and an external torque was applied that turned the suspended system until the restoring torque of the twisted wire was just equal to the external torque so that the system came to rest. It may be recalled, however, that Cavendish determined the magnitude of the restoring torque for a given twist by studying the free oscillations of his suspended system after an initial displacement from equilibrium, in other words, by using his system as a torsion pendulum. It is an oscillating system of this type that we wish to discuss.

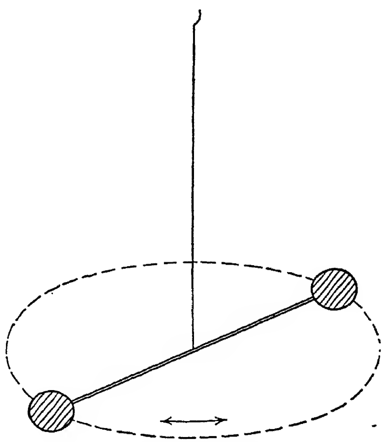


FIG. 219.—Torsion pendulum. Twisting is resisted by the elasticity of the suspension which gives a restoring torque proportional to the angular displacement from equilibrium.

14. It may be recalled that in connection with the discussion of Hooke's law we introduced the coefficient of rigidity or rigidity

modulus to express the elastic resistance of a material to twisting or shearing strain (Chap. II, Par. 32). The application of the general definition given then to the particular case of a twisted wire requires the use of integral calculus. We shall simply quote the result; namely, the torque required to twist the end of a wire of radius r and length l through an angle ϕ is given by

$$L = \frac{\pi}{2}\frac{nr^4}{l}\phi \tag{10}$$

where n is the rigidity modulus. If n is given in newtons per square meter as in the table in Table 1, page 40, r and l in meters and ϕ in radians, then L is in newton-meters.

15. To give an example, suppose we have a wire of steel ($n = 8.0 \times 10^{10}$ newtons/sq. m.) 1 m. long and 0.1 mm. radius. Then a torque of 10^{-4} newton-m. applied to the end will turn it through an angle ϕ given by

$$10^{-4} = \frac{\pi \times 8 \times 10^{10}(10^{-4})^4}{2}\phi = 4 \times 10^{-6}\pi\phi$$

or

$$\phi = \frac{25}{\pi} \text{ rad.}$$

16. For convenience, we may introduce the so-called torsional rigidity of the wire which we may call R. It is the torque required to produce a twist of 1 rad. in the wire, and from Eq. (10) is seen to be $(\pi/2)(nr^4/l)$. Using it, we can rewrite (10)

$$L = R\phi \tag{11}$$

17. Now we are ready to consider the oscillatory motion of a torsion pendulum. Suppose a torsion pendulum is twisted through an angle ϕ and then the external torque is suddenly removed. The restoring torque of the wire will cause an angular acceleration toward the equilibrium position. The angular momentum thus imparted will cause the pendulum to twist through the equilibrium position. The direction of the elastic torque reverses as the system passes through equilibrium, gradually slows down the motion, and reverses it. This oscillation goes on continually until the frictional forces finally bring the system

to rest. If the angular displacement from equilibrium at any time is ϕ, then the restoring torque at that time is

$$L = -R\phi \tag{12}$$

or, in words, the restoring torque is proportional to the angular displacement. The pendulum will, therefore, execute simple harmonic motion since the foregoing condition is equivalent, in rotational motion, to the condition in linear motion that the restoring force be proportional to the displacement. If I is the moment of inertia of the system, then the angular acceleration at any time is

$$\alpha = \frac{L}{I} = -\left(\frac{R}{I}\right)\phi \tag{13}$$

or

$$\frac{d^2\phi}{dt^2} = -\left(\frac{R}{I}\right)\phi. \tag{14}$$

This is of the same form as Eq. (1), Par. 5, and therefore we need only rewrite the relations already derived for linear simple harmonic motion. Substituting ϕ for x, ϕ_0 for c, and R/I for k/m in Eq. (5) we get, for the torsion pendulum,

$$\left.\begin{array}{lrcl}
\text{Torque} & L &=& -R\phi \\
\text{Angular displacement} & \phi &=& \phi_0 \cos\sqrt{\frac{R}{I}}\,t \\
\text{Angular velocity} & \omega &=& -\phi_0\sqrt{\frac{R}{I}}\sin\sqrt{\frac{R}{I}}\,t \\
\text{Angular acceleration} & \alpha &=& -\phi_0\frac{R}{I}\cos\sqrt{\frac{R}{I}}\,t \\
\text{Period} & T &=& 2\pi\sqrt{\frac{I}{R}}.
\end{array}\right\} \tag{15}$$

From the last equation, we see that the torsional rigidity of the wire can be determined by measuring the period of oscillation of a torsion pendulum bob of known or calculable moment of inertia. This is what Cavendish did in the experiment described in Chap. IV, Par. 10.

The Simple Pendulum

18. The simple pendulum is one of the most familiar of all oscillating systems. It consists of a small comparatively heavy body, called the pendulum bob, swinging at the end of a light string or wire from a fixed support. As long as the arc through which the bob swings is very small in comparison with the length of the pendulum, the motion is very nearly simple harmonic motion.

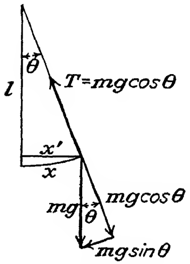


FIG. 220.—The forces acting on a simple pendulum. The motion is simple harmonic if θ is so small that $x = x'$.

19. Clearly the pendulum is in equilibrium when it is hanging vertically. Its displacement from this position can be measured by the length of the arc x in Fig. 220. Therefore, for simple harmonic motion, the restoring force on the pendulum must be proportional to x. The only forces acting on the pendulum bob are its weight and the tension in the string. Resolve the weight into components parallel and perpendicular to the string. The parallel component $mg \cos \theta$ is counteracted by the tension in the string. The resultant force, therefore, is $F = -mg \sin \theta$ tangent to x. If this is proportional to x, we have simple harmonic motion. But $\sin \theta = \frac{x'}{l}$ so that $F = \left(\frac{-x'}{l}\right)(mg)$. Now if θ is very small, x and x' are approximately equal so that we have $F = -x(mg/l)$ and the condition for simple harmonic motion is satisfied. By comparison with Eq. (1), Par. 5, we see that k, the constant of proportion, is, in the present case, equal to mg/l. We can write Eqs. (5) in the form appropriate to the simple pendulum. They become

$$\left.\begin{aligned} T &= 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{l}{g}} \\ \text{Displacement } x &= c \cos \sqrt{\frac{g}{l}}\,t \\ \text{Velocity } v &= -c\sqrt{\frac{g}{l}} \sin \sqrt{\frac{g}{l}}\,t \\ \text{Acceleration } a &= -c\frac{g}{l} \cos \sqrt{\frac{g}{l}}\,t, \end{aligned}\right\} \qquad (16)$$

showing that the period of a simple pendulum is independent of the mass of the bob and the amplitude of oscillation (so long as it is small) but depends on the square root of the length. Evidently, if the amplitude of oscillation is large, we can no longer set $x = x'$ in the preceding argument, the motion is then not simple harmonic motion, and Eqs. (16) do not apply.

Damping

20. In treating various examples of simple harmonic motion, we have neglected the effect of friction. It can be shown both theoretically and experimentally that as long as the friction is small its only effect is to reduce the amplitude of successive oscillations. The period and general behavior of the oscillation are unaffected. Frictional effects may occur in the oscillating system, as in the bearings of a pendulum or in the internal friction of a coiled spring, or may result from the viscous resistance of the medium in which the system is moving. If this medium is air, the effects of viscosity are small and the change of amplitude in an oscillation will only be noticeable after several oscillations. It may take several hours for the swings of a long simple pendulum to die out entirely. If the oscillation is in some viscous medium like water or glycerin, the successive amplitudes diminish rapidly, and in an extreme case, the viscous resistance may be so great that no oscillation occurs at all. The reduction of the amplitude of an oscillation by friction is called damping. If the reduction is great, the oscillation is said to be heavily damped. Any oscillation where the effect is appreciable is said to be a damped oscillation. In Fig. 221, the graph of displacement vs. time is given for such an oscillation.

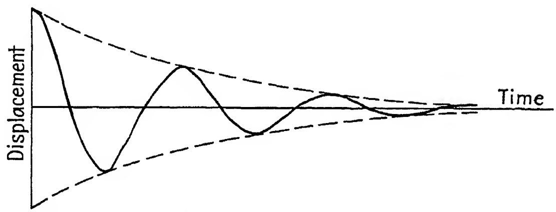


FIG. 221.—Graph of displacement vs. time for a heavily damped oscillation.

Composition of Simple Harmonic Motions

21. We have seen in Par. 10 that a simple harmonic motion can be represented by a vector rotating at constant angular velocity. Evidently this idea is easily extended to any periodically varying quantity which can be represented by a sine or cosine function. In general, such a quantity will be represented by $c \cos (\omega t + \phi)$. For example, in simple harmonic motion not only the displacement can be represented by such a vector whose length is proportional to the amplitude and which rotates

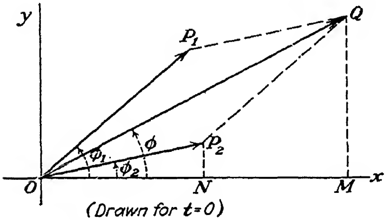


FIG. 222.—Composition of any two simple harmonic motions of the same period. The two simple harmonic motions are represented by the rotating vectors OP_1 and OP_2 and the resultant simple harmonic motion by the rotating vector OQ.

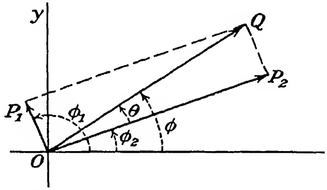


FIG. 223.—Composition of two simple harmonic motions of the same period and differing in phase by exactly 90°.

with an angular velocity $\omega = 2\pi$, but the velocity and acceleration can also be represented by rotating vectors of lengths $c\omega$ and $c\omega^2$, respectively, rotating with the same angular velocity, but with different phases. Similarly, a periodic force applied to a mechanical system or a periodic e.m.f. applied to an electrical circuit can be represented by a rotating vector. Such a vector representation is extremely convenient if we are considering the effect of two periodic vibrations occurring simultaneously.

22. Suppose that a given particle is executing two different simple harmonic motions simultaneously. The two motions have different amplitudes and different phases but the same frequency. Then the two separate motions can be represented by the two rotating vectors OP_1 and OP_2 in Fig. 222, the lengths of the vectors representing the amplitudes, and the angles between the vectors and the x-axis representing the phases. From our previous knowledge of vectors, we might expect that the combined effect of the two motions would be that represented by the vector OQ

formed according to the familiar parallelogram law for vector addition. This expectation is fulfilled. The resultant motion of the particle is therefore a simple harmonic motion with the same frequency as the two component motions, with an amplitude given by OQ, and with a phase angle given by ϕ. In case $\phi_1 - \phi_2$ is 90° (Fig. 223), the resultant amplitude is simply the square root of the sum of the squares of the two component amplitudes and the phase is $\phi_2 + \theta$ where $\theta = \arctan OP_2/OP_1$.

23. For the benefit of the skeptical who feel that the procedure just described is a dubious extension of the laws of vector addition, we shall include an analytical discussion of the problem. Furthermore, such a discussion can obtain exact algebraic results for the amplitude and phase even when the two components of the motion do not differ in phase by exactly a right angle.

The total displacement of the particle at any time t is

$$x = a_1 \cos (\omega t + \phi_1) + a_2 \cos (\omega t + \phi_2) \tag{17}$$

where $a_1 = OP_1$ and $a_2 = OP_2$ are the amplitudes and ϕ_1 and ϕ_2 are the epochs of the two simultaneous simple harmonic motions. Expanding, we obtain

$$\begin{aligned} x &= a_1 \cos \omega t \cos \phi_1 - a_1 \sin \omega t \sin \phi_1 + a_2 \cos \omega t \cos \phi_2 - a_2 \sin \omega t \sin \phi_2 \\ &= (a_1 \cos \phi_1 + a_2 \cos \phi_2) \cos \omega t - (a_1 \sin \phi_1 + a_2 \sin \phi_2) \sin \omega t. \end{aligned} \tag{18}$$

Write

$$a_1 \cos \phi_1 + a_2 \cos \phi_2 = R \cos \phi$$

and

$$a_1 \sin \phi_1 + a_2 \sin \phi_2 = R \sin \phi. \tag{19}$$

Then

$$x = R \cos \phi \cos \omega t - R \sin \phi \sin \omega t = R \cos (\omega t + \phi) \tag{20}$$

which is a periodic displacement with the same frequency as the original motions but with different amplitude and phase. To get the amplitude R and phase ϕ in terms of the original amplitudes and phases, we proceed as follows. Square Eqs. (19), and add them. Then

$$\begin{aligned} R^2(\sin^2 \phi + \cos^2 \phi) = a_1^2(\sin^2 \phi_1 + \cos^2 \phi_1) + a_2^2(\sin^2 \phi_2 + \cos^2 \phi_2) \\ + 2a_1a_2 \cos \phi_1 \cos \phi_2 + 2a_1a_2 \sin \phi_1 \sin \phi_2 \end{aligned}$$

$$R = a_1^2 + a_2^2 + 2a_1a_2 \cos (\phi_1 - \phi_2). \tag{21}$$

To find ϕ, divide $R \sin \phi$ by $R \cos \phi$ giving

$$\tan \phi = \frac{a_1 \sin_1 \phi_1 + a_2 \sin_2 \phi_2}{a_1 \cos_1 \phi_1 + a_2 \cos_2 \phi_2} \tag{22}$$

We have yet to prove that these results are consistent with our vector addition. In the triangle OP_2Q, we know from trigonometry that

$$OQ^2 = OP_2^2 + P_2Q^2 - 2OP_2 \times P_2Q \cos \angle OP_2Q$$

but $\angle OP_2Q = \pi - P_1OP_2 = \pi - (\phi_1 - \phi_2)$; therefore,

$$\cos \angle OP_2Q = -\cos (\phi_1 - \phi_2).$$

Of course $OP_2 = a_2$ and $P_2Q = a_1$ and $OQ = R$ by definition; therefore,

$$R^2 = a_1^2 + a_2^2 + 2a_1a_2 \cos (\phi_1 - \phi_2)$$

in agreement with our analytical result.

To get the phase angle from the geometry of Fig. 222, we have

$$ON = a_2 \cos \phi_2$$

and $NM = a_1 \cos \phi_1$ so that $OM = a_2 \cos \phi_2 + a_1 \cos \phi_1$. Similarly, $MQ = a_1 \sin \phi_1 + a_2 \sin \phi_2$ so that

$$\tan \phi = \frac{QM}{OM} = \frac{a_1 \sin_1 \phi_1 + a_2 \sin_2 \phi_2}{a_1 \cos_1 \phi_1 + a_2 \cos_2 \phi_2}.$$

Forced Oscillations and Resonance

24. The oscillations we have been discussing all resulted from a single initial displacement from equilibrium. An external force was applied and work done in producing this initial displacement, but once the oscillation started there was no external force applied (except damping). The energy put into the system by the initial displacement was the only energy available for maintaining the oscillations. Originally all potential energy, it changes to kinetic and back to potential with each swing of the system and is gradually used up by the damping forces. But we are familiar with oscillations that are maintained by outside sources of power, the motion of the pistons of a steam or internal-combustion engine, for example, or the pendulum of a clock. (We need not be bothered by the fact that the motion of an ordinary piston is not simple harmonic motion.) The piston of a steam engine has no forces on it tending to bring it to any particular position. If there is no steam coming into the cylinder, there is no force on it, and it will be in neutral equilibrium in any position. Therefore we cannot set up in it the kind of natural oscillation we have been talking about. The oscillatory motion it undergoes under the action of steam is produced by a valve system which feeds in

steam alternately on one side of the piston and on the other. The oscillating motion so produced can be of any period depending on the steam pressure and the arrangement of valves. In other words, we have a forced oscillation whose period depends entirely on the period of force producing the oscillation. Simple harmonic oscillations of this type can be produced by a connecting rod coupled to a flywheel in the manner shown in Fig. 224. The period of oscillation of the end of the connecting rod depends only on the speed of rotation of the wheel.

25. The oscillations of a clock pendulum are also forced oscillations but of quite a different sort. Here we are dealing with a

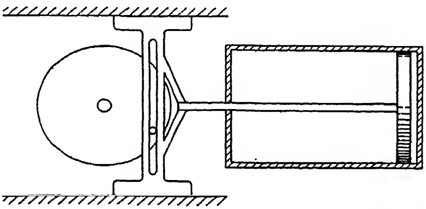


FIG. 224.—An arrangement for giving simple harmonic motion to a piston by the uniform rotation of a flywheel.

system that has a definite position of stable equilibrium toward which a force is continually trying to return it. In addition to this restoring force, periodic impulses are given to the pendulum by the escapement of the clock. These impulses are very small and automatically have the same period as the pendulum since the escapement is released by the pendulum at a certain point of its swing. These small forces applied with the same period as that of the pendulum serve to keep it going with constant amplitude in spite of the damping effect of friction. The effect of the pendulum is to control the escapement, which in turn controls the rotation of the hands so that the rate of rotation of the hands depends ultimately on the natural period of the pendulum.

26. Although it would be possible to swing the pendulum with a frequency different from its natural period by applying a periodic force of sufficient strength, the amplitude obtained would be very much smaller per unit of applied force. Any body that has a natural period of vibration will respond very readily to an applied force that has this same period. This phenomenon is

called resonance and is one that is quite familiar in ordinary life. As examples we may mention "pumping" on a swing, periodic pushing on an automobile stuck in a hole, or the classical example of men marching on a bridge whose natural period is the same as the beat of the march so that the bridge begins to sway dangerously. Resonance can be shown simply with a series of light steel rods of different lengths, each carrying a weight at its end. These rods form pendulums of different periods and are all mounted on the same rigid support. By oscillating the support with different periods, the rods can be made to vibrate, but the vibration will be very small except for the rod that has the same period as the imposed force.

27. As a final example of forced mechanical oscillations, consider a car on a level track. A very light car with frictionless bearings can be pushed back and forth easily, and the direction of motion of the car will reverse the moment the direction of the applied force is reversed. But if the car is heavy, its inertia will always tend to keep it moving in the same direction so that it will not reverse until some time after the direction of the force on it has reversed. We describe this technically as a difference in phase between the applied force and the motion and say that the motion lags behind the phase of the applied force. Finally, suppose that the car is held in a position of equilibrium by springs and that there is friction. Then for any displaced position there are two forces acting, the applied force and the restoring force of the springs, and as soon as there is any motion, there is frictional resistance in addition. The mathematical treatment of this case is somewhat beyond us. We shall merely quote the result that the motion will have the period of the applied force but that the amplitude of that motion for a given amplitude of the applied force will depend on the relation between the period of the applied force and the natural period of the oscillating system. We have already seen that when the two periods are equal we have "resonance," *i.e.*, a very large amplitude produced by a very small force.

Electric Oscillations

28. An important type of motion in physics is the motion of electric charges, *i.e.*, electric currents. In Chap. XVII, we dealt with the laws of currents flowing in response to constant e.m.fs.

But in Chap. XXII on Induced Currents, we saw that the simplest type of generator produces a variable e.m.f. Moreover, it is common knowledge that "alternating currents" are frequently encountered. Therefore it behooves us to investigate the oscillatory motion of electric charges and the effect of alternating e.m.fs. in producing such oscillatory currents. First, we shall discuss free oscillations and then, principally in the next chapter, forced oscillations or alternating currents.

29. Perhaps an electrical oscillation can be understood more easily if we first compare the quantities involved with those encountered in mechanical oscillations. What are the electrical analogues of friction, inertia, and restoring force? Friction always opposes the motion and continually turns kinetic energy into heat. Its analogue in electricity is electrical resistance. Inertia does not resist motion; it resists any *change in motion.* It may act to retard an increase in velocity or a decrease in velocity. We have met a similar effect in electricity, something that tends to react against any *change in current,* namely, self-inductance. To find the electrical analogue to the restoring force, we have to recall some effects discussed earlier. What are we looking for? The effect of the restoring force in a mechanical system is that work is done against it until all the kinetic energy of the moving system has been stored as potential energy. In other words, a reservoir of energy is provided that periodically is emptied when the velocity of the moving system builds up to a maximum and filled when the velocity falls to zero. A similar reservoir of electrical energy is found in a condenser, and the potential difference between the two sides of a condenser is analogous to a restoring force. Evidently a condenser discharging through an inductance forms an electrical system having elements analogous to those in an oscillating mechanical system.

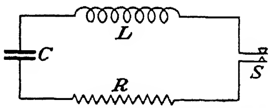


FIG. 225.—Oscillating electric circuit consisting of a capacity, inductance, and resistance in series.

30. Consider the circuit shown in Fig. 225, consisting of a condenser of capacity C, an inductance L, and a resistance R all in series and forming a closed circuit through the switch S. Suppose that initially the switch is open and a charge Q is put on the condenser; this sets up a potential difference $V = Q/C$ between the two sides of the condenser. Work has been done to set up

this potential difference and is represented by the potential energy of the charged condenser. It can be recovered by allowing the condenser to discharge. Suppose we start such a discharge by closing the switch S. Current starts to flow in the circuit but is opposed by the resistance R and the self-inductance L. At any instant, the potential difference V across the condenser must be just equal and opposite to the potential drop in the circuit. This latter consists of two parts RI across the resistance and $L\frac{dI}{dt}$ across the inductance where I is the instantaneous value of the current and dI/dt is its rate of change. As soon as the current reaches a maximum and begins to fall, the inductance acts to maintain it. The result is that the current flows even after the condenser is completely discharged and continues to flow until the condenser is charged sufficiently in the opposite direction to stop it. The condenser is now charged, and the current has stopped; therefore, the system is in exactly the same condition from which it started except for the reversal of the sign of the charges. It will therefore discharge again, the current gradually building up in the opposite direction, and the whole process repeating itself. Were there no resistance in the circuit, this oscillation would continue indefinitely just as the oscillation of a pendulum would continue indefinitely in the absence of friction. The effect of the resistance is to convert some of the energy into heat (I^2R) so that the amount of charge on the condenser diminishes with each oscillation.

31. The general similarity between this electrical oscillation and a mechanical one is obvious. The charged condenser corresponds to a pendulum at the top of its swing. The effect of the inductance in maintaining the current until the condenser is charged in the opposite direction corresponds to the effect of the momentum of the pendulum in carrying it through its position of equilibrium to a displaced position on the other side, and so on. That the analogy is remarkably close may be seen by considering the condenser discharge quantitatively.

32. Since the condenser is the only place in the circuit where charge can accumulate, any motion of charge in the circuit in one direction or the other must result in a change in the charge on the condenser, and vice versa. The rate of change of the charge on the condenser is, then, equal to the current through the circuit, or

in mathematical language, $I = -\frac{dQ}{dt}$ where the minus sign means we are calling the current positive when the condenser is discharging. Also the potential drop through the circuit must equal that across the condenser as we have seen. Writing this out mathematically, we have

$$RI + L\frac{dI}{dt} = V; \tag{23}$$

but we know that $V = Q/C$ and we have just seen that $I = -dQ/dt$. Making these substitutions, we get

$$R\frac{dQ}{dt} + L\frac{d^2Q}{dt^2} = \frac{-Q}{C}. \tag{24}$$

In treating mechanical oscillation mathematically, we neglected friction. Suppose we follow that precedent and call $R = 0$. The equation above then reduces to

$$\frac{d^2Q}{dt^2} = -\frac{1}{LC}Q \tag{25}$$

which we see is exactly similar to Eq. (1) for a mechanical oscillating system. The same mathematical treatment can be applied and results in the following equations, in which Q_0 is the original charge on the condenser,

$$Q = Q_0 \cos\left(\frac{t}{\sqrt{LC}}\right) \tag{26a}$$

$$I = \frac{dQ}{dt} = -\frac{Q_0}{\sqrt{LC}} \sin\left(\frac{t}{\sqrt{LC}}\right) \tag{26b}$$

$$\frac{dI}{dt} = \frac{d^2Q}{dt^2} = -\frac{Q_0}{LC} \cos\left(\frac{t}{\sqrt{LC}}\right) \tag{26c}$$

$$T = 2\pi\sqrt{LC}. \tag{26d}$$

33. Thus we see that a circuit containing inductance and capacity in series has a natural period of oscillation proportional to the square root of the product of the inductance and the capacity. In any real circuit, there will of course be some resistance, but it can be shown that if this is small it modifies the period only slightly. It might be mentioned, however, that the inductances and capacities available for laboratory purposes are such

that it is not possible to construct a circuit having a period of more than about one second. It is possible to get inductances and capacities that give $2\pi\sqrt{LC}$ greater than that, but with any inductances commercially available, the resistance is then so high that the oscillation is too heavily damped to be observable.

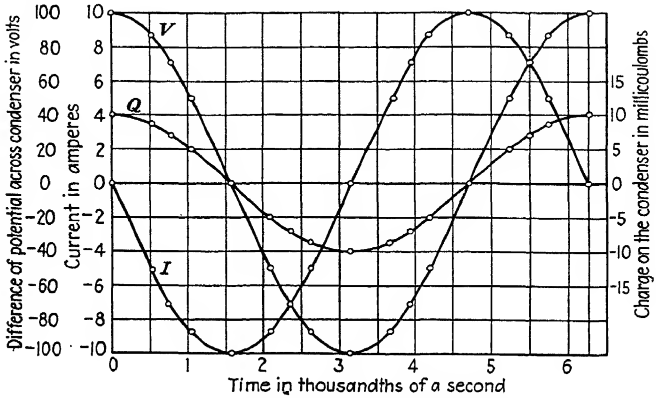


FIG. 226.—Charge Q, current I, and differences of potential V in an oscillatory circuit consisting of 10 millihenries inductance in series with 100 microfarads capacity. $Q = Q_0 \cos (t/\sqrt{LC})$, $I = -(Q_0/\sqrt{LC}) \sin (t/\sqrt{LC})$, and $V = (Q_0/C) \cos (t/\sqrt{LC})$. Condenser raised initially to a potential difference of 100 volts.

The circuits used in radio, which obey these same fundamental principles, have periods that are very much smaller, of the order of a thousandth to a ten-millionth of a second.

Forced Electrical Oscillations

34. We saw that the motion of a mechanical system under the action of a periodic force depended on the relation of the natural period of the system to the period of the force. The same thing is true for an electrical system. If the e.m.f. impressed on the system is periodic, the current through the system will also be periodic and of the same period. But the amplitude or peak value of the current will be much greater if the period of the impressed e.m.f. is the same as the natural period of the circuit. As in mechanical systems, this phenomenon is called resonance. The tuning of a wireless receiver is essentially merely the altering of the inductance or capacity of the circuit until its natural period is that of the sending circuit that is to be heard.

SUMMARY

Simple harmonic motion is the oscillation of a body about a position of equilibrium toward which it is drawn by a restoring force proportional to its displacement from that position. If we start from the assumption $F = -kx$ where F is the restoring force, x the displacement, and k a factor of proportion, the following equations are derived:

$$\text{Displacement } x = c \cos \sqrt{\frac{k}{m}}\,t$$

$$\text{Velocity} \quad v = -c\sqrt{\frac{k}{m}} \sin \sqrt{\frac{k}{m}}\,t$$

$$\text{Acceleration} \quad a = -c\frac{k}{m} \cos \sqrt{\frac{k}{m}}\,t$$

$$\text{Period} \quad T = 2\pi\sqrt{\frac{m}{k}}.$$

The amplitude c is the maximum value of the displacement. The frequency is the number of oscillations per second and is equal to $1/T$. A vector rotating with constant angular velocity represents a simple harmonic motion since the projection of the end of the vector on a diameter of the circle described by that end executes simple harmonic motion.

Examples of simple harmonic motion include a column of liquid oscillating in a U tube, a weight moving up and down on a spring, and a simple pendulum making very small oscillations. A torsion pendulum executes rotational simple harmonic motion. The period of a torsion pendulum is given by $T = 2\pi\sqrt{I/R}$ where I is the moment of inertia of the suspended system and R is the torsional rigidity of the suspension. Thus R can be determined by measuring T.

If friction is appreciable, the amplitude diminishes and the oscillation is said to be a damped oscillation.

If a particle is executing two simple harmonic motions of the same frequency simultaneously, its resultant motion is simple harmonic and represented by a rotating vector which is the vector sum of the two vectors representing the component vibrations.

A periodic force applied to a system having a natural period of its own produces a forced oscillation of the same period as the force. If the period of the force is the same as that of the system, resonance occurs and the amplitude of the oscillation produced will be relatively very large.

If an electrical condenser is allowed to discharge through an inductance, charge will flow back and forth until the energy is all dissipated by the resistance of the circuit, just as the bob of a pendulum will swing back and forth until the energy is dissipated by friction. The natural period of an oscillating circuit is $T = 2\pi\sqrt{LC}$ where L is the inductance and C the capacity of the circuit. A dynamo generating alternating e.m.f. forces an alternating current through a circuit in much the same way that forced mechanical oscillations are produced.

ILLUSTRATIVE PROBLEMS

1. When suspended vertically with no weight on its lower end, a spring is 30 cm. long. A mass of 20 g. attached to the lower end of the spring is found to oscillate about an equilibrium position in which the spring is 50 cm. long. With what frequency does the system oscillate?

Equation (5e), page 471, gives the period $T = 2\pi\sqrt{m/k}$ where $m = 0.02$ kg. = the oscillating mass and $k = F/x$, from Eq. (1), page 470, is the constant of the spring. As we are interested in the magnitude only of k, we omit the minus sign which showed that the direction of F was opposite to that of x. Here, $F = 0.02$ kg. $\times$ 9.8 m./sec.^{2} = 0.196 newton = the force stretching the spring, and x = 50 cm. − 30 cm. = 20 cm. = 0.20 m. = the distance the spring is stretched by the force F. Thus,

$$k = F/x = 0.196 \text{ newton}/0.20 \text{ m.} = 0.98 \text{ newton/m.}$$

The period of vibration is thus

$$T = 2\pi\sqrt{\frac{m}{k}} = 2\pi\sqrt{\frac{0.02 \text{ kg.}}{0.98 \text{ newton/m.}}} = 2\pi\sqrt{\frac{0.0204 \text{ kg.-m.-sec.}^2}{\text{kg.-m.}}}$$

$$T = 2\pi \times 0.143 \text{ sec.} = 0.898 \text{ sec.}$$

As in Par. 7, the frequency ν is the reciprocal of the period or

$$\nu = \frac{1}{T} = \frac{1}{0.898 \text{ sec.}} = 1.11 \text{ vibrations/sec.}$$

2. A simple pendulum has a period of 2 sec. at one point on the earth's surface and 2.01 at a second point. If the value of g at the first point is 9.80 m./sec.^{2}, what will it be at the second point?

The period of a simple pendulum, Eq. (16), page 480, is $T = 2\pi\sqrt{l/g}$ where l is the length of the pendulum and g the acceleration of gravity. At the first point, we have therefore

$$T = 2\text{ sec.} = 2\pi\sqrt{\frac{l}{9.80\text{ m./sec.}^2}}$$

At the second point,

$$T = 2.01\text{ sec.} = 2\pi\sqrt{\frac{l}{g}}$$

Squaring these equations and dividing the first by the second give

$$\frac{4\text{ sec.}^2}{4.04\text{ sec.}^2} = \frac{\frac{l}{9.80}\ \frac{\text{sec.}^2}{\text{m.}}}{l/g} = \frac{g\text{ sec.}^2}{9.8\text{ m.}}$$

Therefore,

$$g = \frac{4}{4.04} \times 9.80\text{ m./sec.}^2 = 9.70\text{ m./sec.}^2$$

3. A torsion pendulum consists of a cylindrical rod 30 cm. long, weighing 3 kg. The rod is supported by a wire at the center with the axis of the rod horizontal. If a torque of 0.25 newton-m. rotates the rod through an angle of $45° = \pi/4$ rad., what will be the period of oscillation if the rod is rotated through a small angle from its position of equilibrium and then released?

From Eq. (11), page 478, the torque is equal to the torsional rigidity of the wire times the angle of displacement or $L = R\varphi$ from which

$$R = \frac{L}{\varphi} = \frac{0.25 \times 4}{\pi}\ \frac{\text{newton-m.}}{\text{rad.}} = \frac{1}{\pi}\ \frac{\text{newton-m.}}{\text{rad.}}$$

The moment of inertia of the rod, Table 16, page 378, is

$$I = \frac{ML^2}{12} = \frac{3\text{ kg.} \times (0.3\text{ m.})^2}{12} = \frac{0.09}{4}\text{ kg.-m.}^2$$

The period of a torsional pendulum, Eq. (15), page 478, is

$$T = 2\pi\sqrt{\frac{I}{R}} = 2\pi\sqrt{\frac{0.09\text{ kg.-m.}^2}{4}\pi\frac{\text{rad.-sec.}^2}{\text{kg.-m.}^2}}$$

$$T = 2\pi\frac{3\sqrt{\pi}}{10 \times 2}\text{ sec.} = 1.67\text{ sec.}$$

4. What is the capacity of a condenser that will give a natural frequency of 2 megacycles/sec. when connected in series with an inductance of 4 microhenries?

The period of a series circuit of inductance L and capacity C is given by Eq. (26*d*), page 489, $T = 2\pi\sqrt{LC}$. The frequency is the reciprocal of the period, whence $\nu = 1/T = 1/2\pi\sqrt{LC}$. Squaring both sides and solving for the capacity C, we obtain

$$\nu^2 = \frac{1}{4\pi^2 LC}$$

$$C = \frac{1}{4\pi^2\nu^2 L} = \frac{1}{4\pi^2\left(2 \times 10^6 \frac{1}{\text{sec.}}\right)^2 4 \times 10^{-6} \text{ henry}}$$

$$C = \frac{1 \text{ amp.-sec.}^2}{64\pi^2 \times 10^6 \text{ volt-sec.}} = 0.001582 \times 10^{-6}\frac{\text{coulomb}}{\text{volt}}$$

$$= 1{,}582 \times 10^{-12} \text{ farad} = 1{,}582 \text{ micromicrofarads.}$$

A henry is defined by the Eq. (7), page 455, $E = L\frac{dI}{dt}$, in which the units are volts = henries amp./sec., whence henries are volt sec./amp.

PROBLEMS

1. A mass of 1 kg. when hung on the end of a spring stretches it 5 cm. If a mass of 5 kg. is hung on the spring and displaced from its position of equilibrium, what will be its period of oscillation?

2. A mass of 50 g. suspended from the bottom of a vertical spring is displaced 15 cm. from its equilibrium position and then released. If the Hooke's law constant for the spring is $k = 0.2$ newton/m., what will be the velocity of the mass when it passes a point 9 cm. from its equilibrium position? What will be its acceleration when 4 cm. from its equilibrium position?

3. The level of mercury oscillating in a U tube moves through the position of equilibrium with a velocity of 36 cm./sec. If the mean length of the mercury is 122.5 cm., what is the amplitude of the oscillation of the level in each side? How much time elapses after the level in one side passes the equilibrium position until it first reaches a distance 4.5 cm. therefrom?

4. A simple pendulum whose length is 100 cm. oscillates at one point on the earth with a period of 2.16 sec. Calculate the acceleration of gravity at this point.

5. A pendulum clock has a period of 1 sec. at a place where $g = 9.80$ m./sec.^{2}. Will it gain or lose, and how much per week, if taken to a place where $g = 9.79$ m./sec.^{2}?

6. A simple pendulum forming part of an electric circuit swings between two magnetic poles. The magnetic field is assumed to be uniform and of strength 10^3 amp. turns/m. in the region where the lowest 10 cm. of the

pendulum swings and zero elsewhere. The length of the pendulum is 100 cm. and it is swinging with an amplitude of 10 cm. Find the induced e.m.f. in the circuit at any instant.

7. The moving coil of a certain D'Arsonval galvanometer oscillates on open circuit with a period of 1 sec. If the moment of inertia of the coil and attached mirror is 0.1 g.-cm.^{2}, and the length and radius of the suspension fiber are 10 cm. and 0.05 mm. respectively, compute the rigidity modulus of the fiber.

8. A copper disk of radius 5 cm. and 2 mm. thick is suspended by means of a steel wire fastened to its center, the plane of the disk being horizontal. The wire is 120 cm. long and has a radius of 0.2 mm. The rigidity modulus of steel is 8.0×10^{10} newtons/m.^{2}, and the density of copper is 8.9 g./cm.^{3}. How many torsional oscillations would such a system make in 3 min.?

9. The moment of inertia of a galvanometer coil is I kg.-m.^{2}, and its natural period of vibration is T sec. If its area is A m.^{2}, it has N turns, and is in a radial magnetic field of H amp. turns/m., show that a current of i amp. produces a deflection given by

$$\theta = \frac{iA\mu_0 HNT^2}{4\pi^2 I}.$$

10. A condenser with a capacity of 1.0 microfarad is connected in series with an inductance of 0.04 henry. What is the natural frequency of oscillation of current in this circuit?

11. What inductance connected in series with a condenser of 50 microfarads will form a circuit whose natural frequency is 60 cycles/sec.?

12. What must be the range of capacities of a variable condenser connected in series with an inductance of 0.4 millihenry in order that the natural frequency of the oscillatory circuit may be varied over the "broadcast band" (550 to 1,500 kc./sec.)?

13. An electrical circuit consisting of an inductance and an air condenser in series oscillates with a period T. The air condenser is filled with transformer oil, and the period of the oscillating circuit is now found to be $\frac{3}{2}T$. What is the relative dielectric constant of the transformer oil?

CHAPTER XXIV

ALTERNATING CURRENTS

1. In Pars. 24 to 27 of the last chapter, we gave a brief discussion of forced mechanical oscillations. We now wish to amplify those ideas in order to introduce the subject of alternating currents, which are essentially forced electrical oscillations. Let us go back to the fundamental principles of mechanics, Newton's laws. According to Newton's second law, the rate of change of momentum of a moving body is equal to the applied force, or in the form we have found most useful, $F = ma$. In this law, we can interpret ma as the force of inertia acting in opposition to the applied force and from this point of view can say that the sum of all the forces acting is always zero. Whether we insist on this point of view or not, we can certainly write the equation $F = ma$ in the form

$$F - ma = 0 \tag{1}$$

which we shall find appropriate for the present discussion.

2. Now consider the problem of a car moving on a level track already mentioned in the last chapter. Suppose that we are applying an external force $F = F_0 \cos \omega t$, that the car has a finite mass M, that it is held in a position of equilibrium by springs, and that there is friction acting which is proportional to the velocity. Measure the displacement from equilibrium by x, as usual, and call the direction to the right of equilibrium the positive direction. From the point of view of Par. 1, we have four forces acting whose vector sum (algebraic sum in this simple case) must be zero, or we can say that the applied force must be just equal and opposite to the other three. Of these, the elastic force is proportional to the displacement but in the opposite direction, the frictional force is proportional to the velocity* but in the opposite

* This type of frictional force is assumed in order to make the treatments of mechanical and of electrical oscillations completely analogous. Actually, the friction for a car on a track would be more nearly independent of velocity (see Chap. V).

direction, and the inertial force is proportional to the acceleration but in the opposite direction. All this can be expressed in the following equation:

$$F - kx - Rv - Ma = 0 \tag{2}$$

where k is the elastic constant of the springs and R is a frictional resistance constant. If we substitute for the velocity and acceleration the first and second derivative of x with respect to t and for F its value as a function of t, Eq. (2) becomes

$$F_0 \cos \omega t - kx - R\frac{dx}{dt} - M\frac{d^2x}{dt^2} = 0 \tag{3}$$

or

$$F_0 \cos \omega t = kx + R\frac{dx}{dt} + M\frac{d^2x}{dt^2}. \tag{4}$$

We are not going to make any attempt to solve this equation in a general way, but we do want to consider the physical situations represented by it.

3. We shall consider the three special cases where one of the reaction forces as represented by one of the terms on the right of Eq. (4) predominates over the other two. Two of these special cases have already been mentioned briefly in Par. 27 of the last chapter.

4. *Case (a).* Suppose the friction is large compared with the inertia and elasticity [Fig. 227(*a*)]. This means that the velocity exactly follows the applied force. The moment the force stops the motion stops, since there is negligible inertia or elastic energy to carry it on. If the force changes direction, the direction of motion changes immediately. In other words, the velocity of the car is exactly in phase with the applied force and consequently can be represented by a cosine function of the same type. In fact we have immediately from Eq. (2), if k and M are negligibly small,

$$v = \frac{F}{R} = \frac{F_0 \cos \omega t}{R}, \tag{5}$$

confirming our physical argument.

5. *Case (b).* For the second special case, suppose the friction and the elastic forces of the spring are negligible [Fig. 227(*b*)].

This corresponds to a very heavy car moving back and forth on a track without friction and with no springs attached. The momentum of the car will now evidently carry it along after the

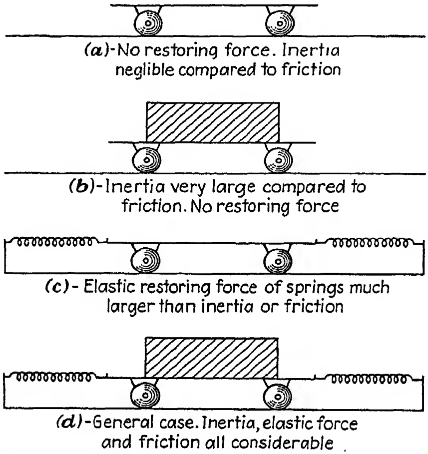


FIG. 227.—Varying conditions which determine the motion of a car on a track in response to an oscillating applied force.

force has reversed. It will take some time for the reversed force to slow the car down and start it moving in the other direction. In other words, the velocity of the car will lag in phase behind the

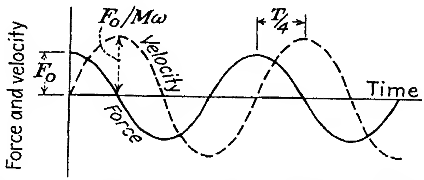


FIG. 228.—Case (b). Friction and restoring force negligible compared to inertia. Graph of applied force and resulting velocity vs. time. Velocity lags a quarter period behind the force.

applied force. We can find the amount of this lag by realizing that the maximum of the velocity must come just when the force changes direction. At this time, the force has been acting continuously in the same direction steadily increasing the velocity, but as soon as the force changes direction, the velocity must begin to decrease. That the velocity must be periodic with the same

period as the force will be evident to the reader if he ponders the impossible consequences of any other assumption. This periodicity being granted, if the velocity is a maximum just when the force is zero and then begins to diminish, it must be lagging behind the force by a quarter of a period, or, in terms of phase, by 90°. This is shown in Fig. 228 where the force and velocity are plotted as functions of the time.

6. These conclusions are confirmed by the mathematical treatment. Equation (2) for this case contains an acceleration term but no velocity or resistance term. To get the velocity, we must use the equation in the form (4) which now reduces to

$$F_0 \cos \omega t = M\frac{d^2x}{dt^2} = M\frac{dv}{dt}. \tag{6}$$

By simple integration which can be verified by students who have studied the calculus, we get*

$$\frac{F_0}{\omega} \sin \omega t = M\frac{dx}{dt} = Mv \tag{7}$$

or

$$v = \frac{F_0}{M\omega} \sin \omega t = \frac{F_0}{M\omega} \cos \left(\omega t - \frac{\pi}{2}\right) \tag{8}$$

which shows that the velocity lags just 90° behind the force in phase although varying periodically with the same frequency. This can also be verified experimentally.

7. *Case (c).* The final special case to consider is that where the friction and inertia are negligible compared with the elastic force of the springs [Fig. 227(*c*)]. The applied force must be always just equal and opposite to the restoring force of the spring since the two other forces opposing it are assumed negligible. Therefore when the displacement is a maximum in the positive direction, the applied force must be a maximum in the positive direction since the restoring force at that time is a maximum toward the position of equilibrium, *i.e.*, in the negative direction. The velocity at this point has decreased to zero and subsequently will increase in the negative direction to a negative maximum when the equilibrium position is reached. During this same

* The constant of integration is omitted since it will be zero if the time is measured from an instant when $v = 0$, *i.e.*, $C = 0$ if $v = 0$ when $t = 0$.

quarter period, the applied force will decrease from a positive maximum to zero at the equilibrium position and then begin to be negative. Thus the applied force is a quarter period, or 90°, behind the velocity in phase. Since we are taking the phase of the applied force as standard, we prefer to say that the velocity leads the applied force in phase by 90°. The velocity and

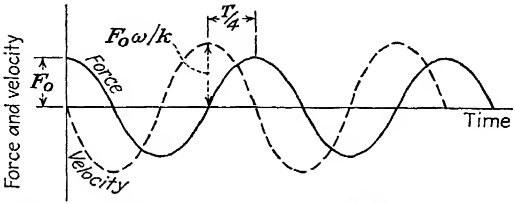


FIG. 229.—Case (c). Friction and inertia negligible compared to restoring force. Graph of applied force vs. resulting velocity. Velocity leads force in phase by 90°.

applied force under steady conditions are plotted in Fig. 229. Mathematically for this case, Eq. (4) reduces to

$$F_0 \cos \omega t = kx \tag{9}$$

from which we can get the velocity by simple differentiation as

$$k\frac{dx}{dt} = kv = -F_0\omega \sin \omega t \tag{10}$$

or

$$v = \frac{F_0\omega}{k} \cos\left(\omega t + \frac{\pi}{2}\right). \tag{11}$$

This confirms directly the result of physical reasoning that the velocity leads the applied force in phase by 90°.

8. We have gone through this rather elaborate discussion of forced mechanical oscillations not so much for its own sake as in the hope of making the analogous problem in the flow of alternating currents more easily understood. Now that we have considered the three special cases of friction only, inertia only, and elasticity only and shown that velocity is in the same phase, 90° behind, and 90° ahead of the force, respectively, in the three cases, we might attack the general case where two or three of these factors are present simultaneously [Fig. 227(*d*)]. But we

believe that this would help us little with the electrical problem. In the simple cases that we have chosen our intuitive notions of velocity and force could contribute to an understanding of the physics of the problem, but this advantage is lost in the greater complexity of the general case. Consequently, the general case will be discussed only for an electrical circuit, but, first, we shall discuss the simplified electrical circuits analogous to the three mechanical situations considered above.

Forced Electrical Oscillations in a Circuit

9. Suppose that we have an a-c generator such as the simple coil described in Par. 30 of Chap. XXII. It produces an alternating e.m.f. given by $E = E_0 \cos \omega t$ where E_0 is the amplitude or maximum value, generally called the "peak voltage" or "peak value" of the e.m.f. Connect this to a circuit such as is shown in Fig. 225 of the last chapter, putting it in where the key S is shown. (For convenience, the circuit diagram is repeated in Fig. 230.) We then have a perfectly general series type of electric circuit containing resistance, inductance, and capacity and a source of e.m.f. This circuit is the exact electrical analogue of the mechanical system described in Par. 2. The e.m.f. of the generator tends to drive electrons through the circuit. This motion is opposed by the resistance, which acts like friction; by the inductance, which acts like inertial mass; and by the difference of potential between the terminals of the condenser, which acts like a spring. Just as the total force on the mechanical system was zero, the total e.m.f. added up around the whole electrical circuit must be zero. In order to apply this principle correctly, we must be clear as to which is the positive direction in the circuit and which is the negative. Let us assume that in Fig. 230 all e.m.fs. in a counterclockwise direction are positive and all e.m.fs. in a clockwise direction are negative. Thus in half of its cycle, the alternating applied e.m.f. will be positive, sending current around the circuit in a counterclockwise direction. Current moving in this direction will flow into the condenser. There will be a potential difference built up in the condenser proportional to its charge and tending to

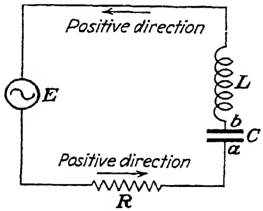


FIG. 230.—A resistance, a capacity, and an inductance in series with a source of alternating e.m.f.

diminish that charge, *i.e.*, a negative e.m.f. according to our convention so long as the charge on the side a of the condenser is positive. We find it convenient to call such potential differences or e.m.fs. built up by the current in various parts of the circuit "back e.m.fs." Thus the resistance will also oppose the flow of current, and this opposition is expressed by calling the product RI a back e.m.f., always in the direction opposite to the flow of current, *i.e.*, negative if I is positive. Similarly, the inductance sets up a back e.m.f. proportional to the rate of change of current and therefore negative if positive current is increasing. At any instant, the algebraic sum of these back e.m.fs. must give a total back e.m.f. equal and opposite to the applied e.m.f. (In the special case of d-c currents, this general law reduces to Kirchhoff's second law in the form given in Chap. XVII, Par. 8.) The phase relations between the different components of the resultant back e.m.f. will be discussed presently. The equations expressing the principles we have just enunciated are analogous to (3) and (4), and from Par. 32 of the last chapter are seen to be

$$E - \frac{Q}{C} - RI - L\frac{dI}{dt} = 0 \tag{12}$$

$$E_0 \cos \omega t = \frac{Q}{C} + R\frac{dQ}{dt} + L\frac{d^2Q}{dt^2}. \tag{13}$$

Although the form of these equations is exactly the same as that of (3) and (4), we should notice that the reciprocal of the capacity corresponds to the elastic constant in the mechanical equations. Thus the analogy to no elastic restoring force is infinite capacity, not zero capacity. As in the case of the mechanical equation, we shall consider the three special cases where one of the right-hand terms of (13) predominates over the other two.

10. *Case* (*a*). If the electrical resistance term is large compared with the two other terms, *i.e.*, the inductance is negligible and the capacity very large or short-circuited, then we simply have

$$E_0 \cos \omega t = R\frac{dQ}{dt} \tag{14}$$

or

$$E = RI \tag{15}$$

which is Ohm's law. Evidently the current and the applied e.m.f. are exactly in phase. The circuit is equivalent to the simple one shown in Fig. 231.

11. *Case* (*b*). As the second type of simplified circuit, consider one with inductance only as shown in Fig. 232. Since we are dealing with a continually alternating applied e.m.f., the current in the circuit is continually changing in response to it. As the e.m.f. of the generator rises from zero, the current in the circuit begins to rise but this sets up an opposing e.m.f. in the inductance which retards the rise in the current. This back

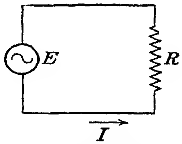


FIG. 231.—A simplified circuit containing only a resistance and a source of alternating e.m.f.

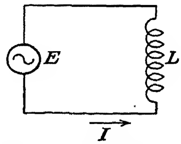


FIG. 232.—A simplified circuit containing only an inductance and a source of alternating e.m.f.

e.m.f. must always be just equal in magnitude to the applied e.m.f.; it is the reaction of the circuit. It is proportional to the rate of change of the current and must change sign when the rate of change of the current changes. As the current rises to a positive maximum the back e.m.f. is negative opposing the current until the maximum is reached when it passes through zero becoming positive to maintain the current and reaching a positive maximum as the current falls to zero. This means that the back e.m.f. lags behind the current in phase by 90°. But the back e.m.f. is equal and opposite to the applied e.m.f., *i.e.*, lags behind it by 180° in phase. Therefore the induced current not only lags behind the applied e.m.f., but lags behind by exactly a quarter period, or 90°, in phase.

12. By turning to Eq. (13), the same result emerges in more clean-cut form. This equation now reduces to

$$E_0 \cos \omega t = L\frac{d^2Q}{dt^2} = L\frac{dI}{dt} \tag{16}$$

which is of exactly the same form as (6) for the mechanical case. Again it can be solved by simple integration which can be verified by students familiar with the calculus. The solution is

$$\frac{E_0 \sin \omega t}{\omega L} = \frac{E_0}{\omega L} \cos\left(\omega t - \frac{\pi}{2}\right) \tag{17}$$

which confirms our conclusion based on physical reasoning that the current lags behind the applied e.m.f. The relation of the current and e.m.f. under steady conditions is shown by Fig. 233. These periodically varying e.m.fs. and currents with which we are dealing can be treated advantageously by the rotating vector method which has already been discussed in the last chapter.

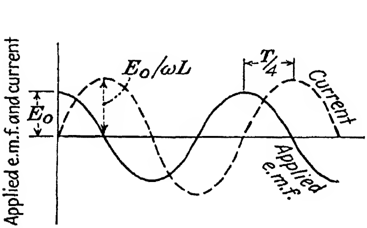


FIG. 233.—Graph of applied e.m.f. and current vs. time in a purely inductive circuit like that of Fig. 232. Circuit has been running long enough to establish a steady state. The current lags behind the e.m.f. in phase by 90° or a quarter period.

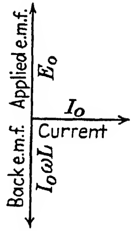


FIG. 234.—Vector representation of the phases of the applied e.m.f., back e.m.f. and current in a purely inductive circuit. The vectors represent amplitudes and are to be thought of as rotating counterclockwise. Their projections on a horizontal line at any instant represent the instantaneous values of the e.m.f., back e.m.f., and current (*cf.* Fig. 233).

This method is particularly useful in studying the phase relations between various quantities. The applied e.m.f., the current, and the back e.m.f. for a purely inductive circuit are shown as rotating vectors with the proper phase relations in Fig. 234. It may be recalled that the counterclockwise direction of rotation is positive. Figures 233 and 234 are merely different ways of representing the same physical result.

13. *Case* (*c*). The last special case we wish to consider is that of a circuit containing only capacity such as is shown schematically in Fig. 235. As usual, we start from the principle that the back e.m.f. must always be just equal and opposite to the applied e.m.f. The back e.m.f. in this case is simply the potential on the condenser $V = Q/C$. As charge flows in the positive direction to the condenser, a back e.m.f. begins to build up opposing

this flow, *i.e.*, a negative e.m.f. This negative e.m.f. has its greatest value when the condenser is fully charged and the current has just sunk to zero. As the current reverses and becomes negative, the charge on the condenser begins to diminish, but the back e.m.f. remains negative becoming zero as the charge falls to zero. The negative current continues to flow, building up a reverse charge on the condenser which gives a back e.m.f. in the positive direction, reaching a maximum when the current changes back from negative to positive. These relations are plotted in Fig. 236, which makes it clear that the back e.m.f. leads the current by 90°, or a quarter cycle. Since the back e.m.f. is 180° out of phase with the applied e.m.f., the current evidently leads the applied e.m.f. by 90° F. The vector diagram in Fig. 237 shows the phase relations in a different way.

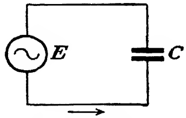


FIG. 235.—A simplified circuit containing only a capacity and a source of alternating e.m.f.

14. This conclusion can be verified immediately mathematically. If the resistance and inductance are negligibly small, Eq. (13) reduces to

$$E_0 \cos \omega t = \frac{Q}{C}. \tag{18}$$

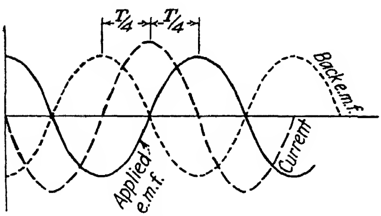


FIG. 236.—Graph of applied e.m.f., back e.m.f. and current vs. time for a steady state in a purely capacitative circuit like that of Fig. 235. The current leads the applied e.m.f. in phase by 90° or a quarter period. The back e.m.f. must by definition lag behind the applied e.m.f. by 180°.

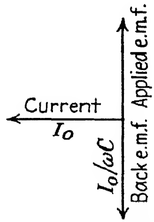


FIG. 237.—Vector representation of the amplitudes and phases of the applied e.m.f., back e.m.f., and current in a purely capacitative circuit (*cf.* Fig. 234 and Fig. 236).

Since the current is the rate of change of the charge Q on the condenser, we can obtain an expression for it by simple differentiation with respect to time; this gives us

$$\frac{dQ}{dt} = I = -\omega C E_0 \sin \omega t \tag{19}$$

or

$$I = \omega C E_0 \cos\left(\omega t + \frac{\pi}{2}\right) \tag{20}$$

which not only confirms our previous conclusion about the phase of the current but gives us an exact expression for it in terms of the constants of the applied e.m.f. and the capacity of the condenser. The relation between the applied e.m.f. and the current over the whole cycle after steady conditions have been established is shown in Figs. 236 and 237.

15. In our mechanical examples and special circuits, the problem was approached from the point of view of an applied force and its effect in producing motion. In every case, the motion or the current was found to be periodic and to have the same frequency as the applied force or e.m.f. This is a general result that is still true for a mechanical system containing friction, inertia, and elasticity simultaneously or for an electric circuit containing resistance, inductance, and capacity simultaneously. This can be verified by solving Eq. (3) or (12) in its general form. We shall make no attempt to do this but shall accept the result as sufficiently established and go on to the discussion of the general electrical case.

16. Suppose we have a periodic e.m.f. applied to a circuit containing resistance, inductance, and capacity in series. Then there will be an alternating current flowing in all parts of the circuit. Furthermore, this current must be continuous; it must have the same value at every point in the circuit at any given time. If this were not so, if the current were to be greater at one point than at another or to flow in one direction at one point and in the reverse direction at another, there would be accumulations of charge which would introduce large back e.m.fs. to smooth out the current again. The only place where charge can accumulate is in the capacity, and this has been taken care of in our equations. We do not concern ourselves with the current distribution in the plates of the condenser itself. The current, then, has the same instantaneous value everywhere. But as we have seen, the back e.m.fs. will vary in both amplitude and phase in different parts of the circuit, depending on whether they arise from resistance, inductance, or capacity. The algebraic sum of the back e.m.fs.

in the various parts of the circuit must always be equal and opposite to the applied e.m.f. This applied e.m.f. is, so to speak, used up in the various parts of the circuit to just the extent necessary to keep the same current flowing through that particular part of the circuit that is flowing in every other part.

17. Let us start by expressing the current as a cosine function of time as follows:

$$I = I_0 \cos \omega t \tag{21}$$

where I_0 is the maximum or peak value of the current and we are arbitrarily setting the phase of the current as initially zero although we shall later express it relative to the phase of the applied e.m.f.

18. In Pars. 10 to 14, we deduced the relations between the phases of the current and the applied e.m.f. for the three special cases of circuits containing resistance only, inductance only, and capacity only. Recapitulating these results in tabular form, we have

Pure resistance................	Current is in phase with applied e.m.f.
Pure inductance...............	Current lags behind applied e.m.f. by 90°
Pure capacity..................	Current leads applied e.m.f. by 90°

19. We wish now to express the relations of the *back* e.m.fs. to the current in various parts of the circuit. This can be done at once if we remember that the back e.m.f. is always 180° behind the applied e.m.f. We have therefore

In a resistance..........	Back e.m.f. lags behind the current by 180°
In an inductance........	Back e.m.f. lags behind the current by 90°
In a capacity...........	Back e.m.f. leads the current by 90° (equivalent to a lag of 270°)

20. These results have already been deduced in Pars. 10 to 14 and the cases of inductance and capacity illustrated in Figs. 233 to 237. They can also be expressed in the following equations based on the phase of the current as standard:

$$E'_R = RI_0 \cos (\omega t - \pi) \tag{22}$$

$$E'_L = \omega L I_0 \cos \left(\omega t - \frac{\pi}{2}\right) \tag{23}$$

$$E'_C = \frac{I_0}{\omega C} \cos \left(\omega t + \frac{\pi}{2}\right). \tag{24}$$

The primed quantities on the left are the back e.m.fs., a notation which we shall use consistently. The amplitudes on the right can be verified by comparison with the equations in Pars. 10 to 14, and the phases are simply the mathematical expression of the statements made in Par. 19. The vectors representing these e.m.fs. are shown in Fig. 238. Furthermore, at the end of the last chapter we saw how to get the resultant effect of any two such vectors. We can very easily get the resultant of three by first combining two and then combining that resultant with the third. Here we

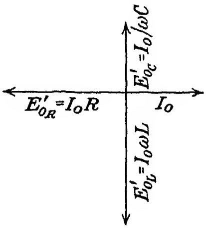


FIG. 238.—Vector representation of the back e.m.f.'s in the resistance, E'_{OR}, inductance, E'_{OL} and capacity, E'_{OC} of a complete series circuit (see Fig. 230) referred to the current, I_0.

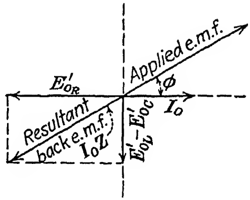


FIG. 239.—Vector diagram of the components of the back e.m.f., the resultant back e.m.f., the current and the applied e.m.f. in a series circuit. This diagram illustrates the graphical determination of the phase difference ϕ between the applied e.m.f. and the current.

have a particularly simple case since all the phase differences are 90° or 180°. We can merely add the inductance and capacity back e.m.fs. algebraically getting a vector $I_0\left(\omega L - \dfrac{1}{\omega C}\right)$ along the vertical axis as shown in Fig. 239. Then this is added vectorially to I_0R which is perpendicular to it. The vector labeled I_0Z in Fig. 239 is the rotating vector representing the resultant back e.m.f. The square of its length is obviously

$$I_0^2Z^2 = I_0^2\left[R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2\right],$$

and therefore

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}. \tag{25}$$

Z is called the *impedance* of the circuit, ωL the *inductive reactance* of the circuit, $1/\omega C$ the *capacitative reactance* of the circuit. The

difference $\left(\omega L - \frac{1}{\omega C}\right)$ is the total *reactance* of the circuit and is usually designated by X.

21. We are now ready to express these results in more customary and more convenient form. The back e.m.f. is always exactly equal and opposite to the applied e.m.f. Therefore the applied e.m.f. has an amplitude I_0Z and differs from the current in phase by the angle ϕ where ϕ is determined by the equation

$$\tan \phi = \frac{E'_{0L} - E'_{0C}}{E'_{0R}} = \frac{\omega L - \frac{1}{\omega C}}{R}.$$

But now let us take the applied e.m.f. as standard and express the magnitude and phase of the current in terms of it. We then have our general equations for the flow of current through a series circuit of resistance, inductance, and capacity in terms of the applied e.m.f. and the constants of the circuit as follows:

If the applied e.m.f. is

$$E = E_0 \cos \omega t, \tag{26}$$

then the current is

$$I = \frac{E_0 \cos (\omega t - \phi)}{Z} \tag{27}$$

where

$$\phi = \arctan \left(\frac{\omega L - \frac{1}{\omega C}}{R} \right) \tag{28}$$

and

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}. \tag{29}$$

The Measurement of Alternating Currents

22. We have said nothing as yet about the frequency of alternating currents and voltages, but it is common knowledge that the frequency of the current in ordinary lighting circuits is 60 cycles per second. Some power circuits run at 25 cycles per second, but we are more likely to encounter higher frequencies than lower. Alternating currents cannot be measured by an

instrument like the D'Arsonval moving-coil galvanometer, the direction of whose deflection depends on the direction of the current. The inertia of the moving coil would be so great it could not possibly follow such rapid changes in direction, and even if it could, our eyes could not observe it. Some instrument must be used which, like the electrodynamometer (see Chap. XVI, Par. 25) or a hot-wire ammeter, is independent of the direction of flow of the current. The heating effect of a current is proportional to the square of the current, and if this current is varying rapidly, the heating effect is proportional to the average value of the square of the current. The square root of this, commonly called the root-mean-square, or r.m.s., current, is the d-c current which flowing steadily would have the same heating effect as the actual a-c current. It is this effective, or r.m.s., current that all a-c instruments are calibrated to measure. For a sinusoidal alternating current, it is related to the peak current as follows:

$$I_{\text{eff}} = \frac{I_0}{\sqrt{2}} = 0.707\, I_0. \tag{30}$$

The effective voltage is then defined as the root mean square of the a-c voltage which will send an alternating current of r.m.s. I_{eff} through 1-ohm resistance. For a sinusoidal e.m.f., it bears the same relation to the peak voltage as above, namely,

$$E_{\text{eff}} = \frac{E_0}{\sqrt{2}} = 0.707 E_0. \tag{31}$$

By definition, Joule's law for the heat developed by a current holds if the effective current is used. The generalized Ohm's law given in Eq. (27) may evidently be written in terms either of peak currents and voltages or of effective values. Thus

$$I_0 = \frac{E_0}{Z} \qquad \text{and} \qquad I_{\text{eff}} = \frac{E_{\text{eff}}}{Z}. \tag{32}$$

The Power Used in an A-C Circuit

23. In an a-c circuit containing resistance, inductance, and capacity, power may be absorbed in three ways, heating the resistance, establishing the magnetic field of the inductance, and

establishing the electric field of the capacity. The energy in the inductance and capacity is at least partly returned to the circuit when the magnetic and electric fields return to zero. It would be entirely so if some heating of the magnetic material of the inductance and of the dielectric material of the condenser did not take place every time the fields in them were reversed. The energy that goes to heating the resistance is, of course, irrecoverable. Evidently if we had a circuit without resistance, we might have enormous currents flowing without appreciable expenditure of energy. In terms of our knowledge of d-c currents, this sounds

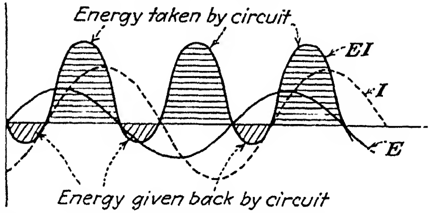


FIG. 240.—Graph of applied e.m.f., current and power vs. time in a circuit where the current is lagging behind the applied e.m.f. by about 50°. The shaded areas above the horizontal time axis represent positive values of the power, *i.e.*, energy supplied to the circuit. The shaded areas below the axis represent negative values of the power, *i.e.*, energy fed back by the circuit.

contradictory but it is not, as we shall see if we actually calculate the power used.

Power Factor

24. To make a calculation of the power used, we must go back to first principles. By definition, in an element of time Δt so small that the e.m.f. and current are constant during it, the amount of work done is $EI\Delta t$. Our problem then is to sum up this product for a whole cycle. This process is done graphically in Fig. 240 which is drawn for a circuit such that the current is lagging about 50° behind the applied e.m.f. It is evident that the product EI of the instantaneous values of E and I is sometimes positive and sometimes negative. The negative value must correspond to the process of feeding back energy to the power supply by the inductance and capacity. If the difference of phase ϕ were 90°, the product EI would be negative just as much as positive and there would be no net consumption of energy, merely a transfer of it back and forth between the circuit and the

power supply. To get $\phi = 90°$, we would have to have no resistance in the circuit which would give the case already mentioned of large flow of current with no expenditure of energy.

25. To calculate the average value of the product EI over a whole cycle requires an integration somewhat beyond us, but the result is so important that we shall quote it. This gives the average energy used up in the circuit per unit time, *i.e.*, the average power supplied to the circuit. It is

$$E_0 I_0 \cos \phi \tag{33}$$

which is commonly written

$$EI \cos \phi \tag{34}$$

where E and I are now the effective, or r.m.s., values of the e.m.f. and current, the values that would be measured by a-c instruments. As usual, ϕ is the difference in phase between applied e.m.f. and current, but Eq. (34) is so important that it gives a name to $\cos \phi$. It is called the power factor. A circuit where the current lags behind the voltage is said to have a lagging power factor, and ϕ is called the angle of lag; if the current leads the voltage, *i.e.*, if ϕ is negative, the circuit is said to have a leading power factor, and ϕ is called the angle of lead.

Transformers

26. We are now in a position to discuss the subject of transformers already touched upon in Pars. 19 and 26 of Chap. XXII. As described there, a transformer consists of two coils insulated from each other but "coupled" so that the alternating magnetic flux set up by the current in one coil induces an e.m.f. in the other coil. The most frequent use of a transformer is to raise or lower the voltage of a power line. The power transmitted by a power line is the product of the current and the difference of potential (power factor for the moment being neglected), but the loss of power in a line is proportional to I^2R. Consequently, for the same product of potential difference and current, a very low current at a high potential will transmit the same amount of power with less heating loss in the line than a large current at a low potential. For transmission, therefore, high voltages are

most economical. But very high voltages are troublesome and dangerous both in generating machinery and in industrial and domestic uses of power. It is frequently desirable to generate power at a low or moderate voltage, to transmit it at a high voltage, and to use it at low voltage. For direct currents, this is not easily done but transformers make it comparatively simple for alternating currents.

27. As we saw in Chap. XXII, the e.m.f. induced in a coil by a changing magnetic flux through it is equal to the product of the rate of change of the flux and the number of turns of the coil. If, therefore, the primary of a transformer is generating a certain alternating magnetic flux most of which is going through the secondary, the potential induced in the secondary depends on the number of turns in it, and by altering the number of turns, the secondary voltage may be made as high or as low as is desired. But if we want to make this secondary voltage do any work, current must flow in the secondary and this may alter the flux through the coils in a way which is not immediately obvious. Suppose we consider the situation more in detail.

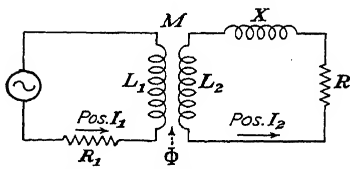


FIG. 241.—Diagram of a transformer circuit. The primary circuit is on the left, the secondary on the right. The resistance R_1 of the primary is considered negligible in our discussion.

28. Let the circuit be as indicated in Fig. 241 where the two inductances L_1 and L_2 are the primary and secondary of a transformer. Let us assume that the coupling is perfect, *i.e.*, that all the magnetic flux set up by current flowing in either coil of the transformer passes through the other coil and that there is a negligible loss of energy in the transformer core even if this core is iron. Suppose that the primary coil has N_1 turns and the secondary coil has N_2 turns. Let M be the mutual inductance of the two coils as defined in Par. 21 of Chap. XXII. Let the alternating e.m.f. $E = E_0 \cos \omega t$ be applied to the primary, and call the current in the primary I_1 and the current in the secondary I_2. The positive direction in the circuits is shown by the arrows in Fig. 241, and it is understood that the coils are so wound that a positive current in the secondary produces a flux in the same direction as that produced by a positive current in the primary.

Then at every instant, the applied e.m.f. in the primary must be just equal to the total back e.m.f. This back e.m.f. consists of three parts, the IR drop from the resistance, the LdI_1/dt from the self-inductance, and MdI_2/dt, the back e.m.f. induced in the primary by the changing current in the secondary. Written as an equation,

$$E = I_1R_1 + L_1\frac{dI_1}{dt} + M\frac{dI_2}{dt}. \tag{35}$$

Similarly, in the secondary, the sum of the induced e.m.f. and the back e.m.f. of the load must be zero since there is no external applied potential. In the form of an equation, this is

$$0 = I_2Z_2 + L_2\frac{dI_2}{dt} + M\frac{dI_1}{dt} \tag{36}$$

where Z_2 is the impedance of the load. We have no intention of attempting to solve these equations as they stand, but we can study them from a physical point of view.

29. Since the transformer in practice is a device for transmitting power, not for consuming it, the resistance of the coils should be as small as possible. In practice, it is so small compared with the inductive reactance that we can neglect it. If the inductive reactance is the only appreciable impedance in the primary circuit, the back e.m.f. from it alone must always equal the applied e.m.f. If Φ is the magnetic flux in the core common to both primary and secondary coils, the back e.m.f. in the primary is $N_1d\Phi/dt$; this, then, must always be equal and opposite to the applied e.m.f. If there is no current in the secondary, the primary current alone must give a back e.m.f. equal to the applied e.m.f. The primary current that has this effect is called the magnetizing current I_μ. If there is current flowing in the secondary, then the primary current will change until the combined effect of primary and secondary current is equivalent to I_μ so that the proper flux and back e.m.f. are produced.

30. Consider first the situation when the secondary circuit is open so that I_2 is zero. Then a current I_μ must flow in the primary. Since the circuit is purely inductive, the current must lag behind the applied e.m.f. by just 90°. By definition, the flux at any instant is proportional to $(N_1I_1 + N_2I_2)$ where I_1 and I_2 are the instantaneous values of the currents and N_1 and N_2

are the numbers of turns in the primary and secondary coils. Since we are taking I_2 as zero, the flux must be in phase with I_1. Taking the phase of the flux as the standard, we can represent the various e.m.fs. involved by vectors as shown in Fig. 242 where E_1 is the applied e.m.f., E_1' the back e.m.f. in the primary, and E_2 the induced e.m.f. in the secondary. If we call the flux Φ, the back e.m.f. from the inductance of the primary is then $N_1 d\Phi/dt$. The induced e.m.f. in the secondary is similarly $N_2 d\Phi/dt$ since we have assumed the coupling so good that the flux in the two coils is always identical. Like the inductive back e.m.f. in the primary, the induced e.m.f. in the secondary is just 90° behind the flux in phase since it is in phase with the rate of change of flux. It is therefore 180° behind the applied primary e.m.f. in phase.

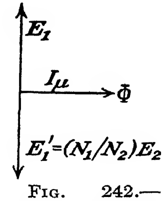


FIG. 242.—Vector diagram of the applied e.m.f. E_1, the current I_μ, and the back e.m.f. E'_1 in the primary of a transformer and the induced e.m.f. E_2 in the secondary when the secondary circuit is open. The flux Φ, here in phase with the current, is taken as the standard of phase in all these transformer diagrams.

NOTE: *In this and subsequent figures* (243–245) *concerning the transformer the subzeros indicating amplitudes, i.e.,* E_0, Φ_0 *etc., have been omitted for simplification.*

31. Now consider what happens when the secondary of the transformer is connected through a load of impedance Z, made up of a reactance X and a resistance R. The induced e.m.f. will cause a current I_2 to flow which will tend to alter the flux through the coils. But practically the only opposition that is offered to the applied e.m.f. in the primary is the back e.m.f. from the alternations of this flux. For a given applied e.m.f., therefore, the amplitude of the flux must remain practically constant. Any changes in the flux, resulting from the flow of current in the secondary, must be immediately counteracted by changes in the primary current. Additional current called the load current I_L must flow in the primary such that the combined effect of the total primary and secondary current must produce a flux identical with that produced by the primary current when the secondary circuit was open. The load current must just counteract the effect of the secondary current on the flux. The resultant primary current and secondary current will not be in phase. In fact, they will be nearly 180° out of phase since the open circuit current I_μ is small compared with the load current in a good power transformer. Therefore they must be added vectorially as

explained in Chap. XXIII, Par. 22. Furthermore, since it is their contribution to Φ that interests us, not the currents themselves, they must be multiplied by N_1 and N_2. We know that their resultant effect must be the same in phase and magnitude as the I_1N_1 on open circuit, *i.e.*, must equal the product $I_\mu N_1$. Also considering the secondary circuit as one in which the induced e.m.f. E_2 is applied to a resistance R and a reactance X in series and assuming the reactance predominantly inductive, we know that I_2 must lag behind the induced e.m.f. E_2. The induced e.m.f. E_2, as we saw in Par. 30, must lag just 90° behind the flux. Incorporating these conclusions in a vector diagram, we get Fig. 243. The flux is again taken as the standard of phase and determines the direction of the resultant current.

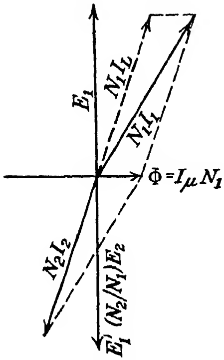


FIG. 243.—Vector diagram for a transformer when the secondary is closed through a resistance and predominantly inductive reactance. The secondary current I_2 lags behind the secondary e.m.f. E_2 which is always 90° behind the flux Φ. Additional load current I_L flows in the primary such that the vector sum of N_1I_L and N_2I_2 equals $I_\mu N_1$.

32. On returning to the primary, our argument has to be altered but little from that for open circuit. The main part of the back e.m.f. is still caused by the alternating flux which is unchanged. Figure 242 is unchanged except that if we wish to interpret the back e.m.f. in terms of currents it is the combined self-inductance effect $L_1I_1\omega$ and mutual inductance effect $MI_2\omega$. This is shown in Fig. 244. A similar vector diagram for the back e.m.f. in the secondary load is given in Fig. 245.

Iron-cored and Air-cored Transformers. Efficiency

33. In the explanation we have given for the operation of a transformer, we assumed that the resistance of the primary was negligible compared with its inductive reactance; we set the inductive back e.m.f. equal to the applied e.m.f. It is evident that this is desirable since any power that goes to overcoming the resistance of the primary is lost as heat and is not transmitted to the secondary. At ordinary power frequencies, it is impossible to wind an inductance coil with an air core that satis-

fies this condition; but by winding the coils on an iron core, the magnetic flux produced by a given current in a given winding is so much increased that the inductive reactance can be made very much larger than the resistance. This is the principal reason for using iron in a transformer. Since the inductive reactance is equal to ωL, it may be so great at high frequencies that an iron core is unnecessary or even undesirable. Another assumption we have made which is not strictly correct was that of perfect coupling. There will always be some loss of energy resulting from

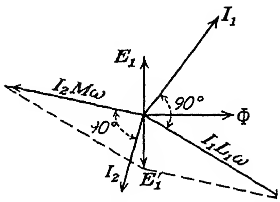


FIG. 244.—Vector diagram showing the back e.m.f. in the primary in terms of the primary and secondary currents and the self and mutual inductance.

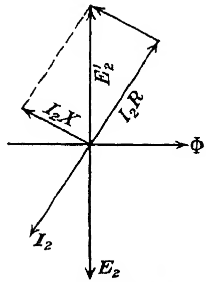


FIG. 245.—Vector diagram showing the back e.m.f., E_2', in the secondary in terms of its components, I_2R and I_2X, in the resistance and reactance of the secondary.

magnetic leakage, the fact that some of the flux produced by the primary does not pass through the secondary, and vice versa. Eddy currents induced in the iron core are another cause of loss. The constant reversal of the direction of magnetization in the iron constitutes another cause of loss of power. The repeated changes of orientation of the elementary magnets in the iron produce heat, dissipating energy which must be supplied by the primary circuit. Finally, there will be some energy lost in the resistance of the secondary windings.

34. All these factors could be taken into account by the methods we have been using. No new principles would be called upon, but the vector diagrams of the circuits would be very much more complicated. Our treatment is justified not only on the grounds of simplification, but because the various losses we have neglected are surprisingly small. The sum total of them all is only 0.5 to 4 per cent of the power transmitted by a good power transformer.

This remarkably high efficiency coupled with the absence of moving parts makes a transformer one of the most satisfactory of machines.

Voltage Ratio and Current Ratio

35. As we saw in Par. 29, the applied e.m.f. in the primary is just equal to the induced back e.m.f. $N_1 d\Phi/dt$ where Φ is the magnetic flux and the induced e.m.f. in the secondary is equal to $N_2 d\Phi/dt$. Therefore the ratio of the e.m.f. induced in the secondary to the applied e.m.f. is simply N_2/N_1, the ratio of the numbers of turns. To get the current ratios is slightly more difficult. In reference to Fig. 243, it is evident that the vector $N_1 I_1$ can be broken up into two vectors, one of which $I_\mu N_1$ corresponds to the magnetizing current that flows on open circuit and the other of which is the product of N_1 and the load current I_L. As we have seen in Par. 31, this load current just counteracts the effect of the secondary current on the flux, therefore $N_1 I_L = N_2 I_2$. In good commercial power transformers, the magnetizing current is only 1 to 10 per cent of the load current so that it is a good approximation to replace the load current in the preceding equation by the total primary current. Rewriting this and the voltage ratio, we have

$$\text{Voltage ratio } \frac{E_1}{E_2} = \frac{N_1}{N_2}. \tag{37}$$

$$\text{Current ratio } \frac{I_1}{I_2} = \frac{N_2}{N_1}. \tag{38}$$

SUMMARY

If a car is moving on a level track under the action of a periodic applied force, the phase relations between the force and the velocity are (*a*) friction large, inertia and restoring force negligible, in phase; (*b*) inertia large, friction and restoring force negligible, velocity lags behind applied force by 90°; (*c*) restoring force large, inertia and friction negligible, velocity leads applied force by 90°.

Similar arguments applied to electric circuits show that (*a*) in a resistance, the current is in phase with the applied e.m.f.; (*b*) in an inductance, it lags behind the applied e.m.f. by 90°; (*c*) in a capacity, the current leads the applied e.m.f. by 90°.

In a circuit containing resistance, inductance, and capacity in series, the relation between applied periodic e.m.f. and the current it produces is obtained by considering the back e.m.fs. in relation to current in the three different parts of the circuit. The result is that if the applied e.m.f. is $E = E_0 \cos \omega t$

$$I = \frac{E_0 \cos (\omega t - \phi)}{Z}$$

where

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \quad \text{and} \quad \phi = \arctan \frac{\left(\omega L - \frac{1}{\omega C}\right)}{R}$$

Z is called the impedance of the circuit and $\left(\omega L - \frac{1}{\omega C}\right)$ the reactance.

Alternating currents are measured in terms of the effective heating current which is the square root of the average of the square of the current. Similarly, the voltage measured in an a-c voltmeter is the root-mean-square voltage. In terms of this voltage and current, the power used in an a-c circuit is $EI \cos \phi$. Cos ϕ is called the power factor.

In the primary circuit of a transformer, the resistance is made negligibly small so that the only back e.m.f. is that of inductance of the transformer windings. To make this high without introducing appreciable resistance, the coil is wound on an iron core unless high frequencies are used. The action of the transformer is followed in detail by using vector diagrams of the various e.m.fs., currents, and the magnetic flux. For small loads on the secondary, the ratio of voltages and currents in the primary and secondary are given by $E_1/E_2 = N_1/N_2$ and $I_1/I_2 = N_2/N_1$ where N_1 and N_2 are the number of turns in the primary and secondary, respectively.

ILLUSTRATIVE PROBLEMS

1. A resistance of 10 ohms is connected in series with an inductance of 30 millihenries. An alternating current with a frequency of 60 cycles per second flows in the circuit. The maximum value of the current is 20 amp. Find the back e.m.f. in the resistance and inductance, and calculate the impedance of the circuit.

The current in the circuit from the data given is

$$I = 20 \text{ amp. } \cos\,[(2\pi \text{ rad.})(60 \text{ sec.}^{-1})t].$$

The back e.m.f. in the resistance is given by Eq. (22).

$$\begin{aligned} E'_R &= RI_0 \cos\,(\omega t - \pi) \\ &= 10 \text{ ohms} \times 20 \text{ amp. } \cos\,[(2\pi \text{ rad.})(60 \text{ sec.}^{-1})t - \pi \text{ rad.}] \\ &= 200 \text{ volts } \cos\,[(2\pi \text{ rad.})(60 \text{ sec.}^{-1})t - \pi \text{ rad.}]. \end{aligned}$$

The back e.m.f. in the inductance is given by Eq. (23).

$$\begin{aligned} E'_L &= \omega L I_0 \cos\left(\omega t - \frac{\pi}{2}\right) \\ &= (2\pi \times 60 \text{ rad. sec.}^{-1})(30 \times 10^{-3} \text{ henry})(20 \text{ amp.}) \times \\ &\qquad \cos\left[(2\pi \times 60 \text{ rad. sec.}^{-1})t - \frac{\pi}{2} \text{ rad.}\right] \\ &= (377 \times 0.6 \text{ sec.}^{-1} \text{ volts-sec.-amp.}) \cos\left[(2\pi \times 60 \text{ rad. sec.}^{-1})t - \frac{\pi}{2} \text{ rad.}\right] \\ &= 226 \text{ volts } \cos\left[(2\pi \times 60 \text{ rad. sec.}^{-1})t - \frac{\pi}{2} \text{ rad.}\right] \end{aligned}$$

The phase relations of the current and back e.m.fs. in the resistance and inductance are shown in Fig. 246.

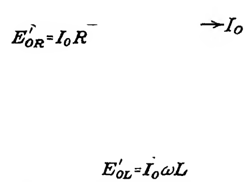


FIG. 246.—Vector diagram of current and back e.m.fs. from a resistance and an inductance in series. Illust. Prob. 1.

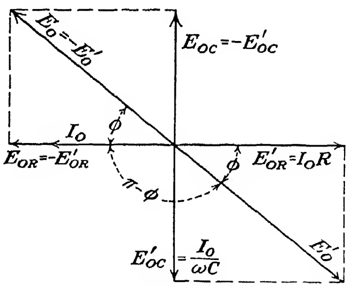


FIG. 247.—Vector diagram of current, back e.m.fs. and applied e.m.f. in a condenser and resistance in series. Illust. Prob. 2. (In this diagram the current vector is toward the left.)

The impedance of the circuit is given by Eq. (29).

$$\begin{aligned} Z &= \sqrt{R^2 + (\omega L)^2} = \{(10 \text{ ohms})^2 + [(2\pi 60 \text{ rad. sec.}^{-1})(30 \times 10^{-3} \text{ henry})]^2\}^{\frac{1}{2}} \\ &= [(10 \text{ ohm})^2 + (11.31 \text{ sec.}^{-1} \text{ volt-sec.-amp.}^{-1})^2]^{\frac{1}{2}} \\ &= [100 \text{ ohm}^2 + 127.9 \text{ ohm}^2]^{\frac{1}{2}} = 15.10 \text{ ohms}. \end{aligned}$$

2. A resistance of 15 ohms is connected in series with a condenser of 200 microfarads capacity, and a current of 30 amp. maximum value and a frequency of 60 cycles per second flows in the circuit. Find the back e.m.f. in (*a*) the resistance, (*b*) the condenser, and (*c*) the entire circuit.

(*a*) The current in the circuit is

$$I = 30 \text{ amp. } \cos\,(2\pi \text{ rad.})(60 \text{ sec.}^{-1})t.$$

The back e.m.f. in the resistance is given as in Prob. 1 by Eq. (22),

$$\begin{aligned} E'_R &= (15 \text{ ohms})(30 \text{ amp.}) \cos\,[(2\pi 60 \text{ rad. sec.}^{-1})t - \pi \text{ rad.}] \\ &= 450 \text{ volts} \cos\,[(2\pi 60 \text{ rad. sec.}^{-1})t - \pi \text{ rad.}]. \end{aligned}$$

(*b*) The back e.m.f. in the condenser is given by Eq. (24).

$$\begin{aligned} E'_C &= \frac{I_0}{\omega C} \cos\left(\omega t + \frac{\pi}{2}\right) \\ &= \frac{30 \text{ amp. } \cos\left[(2\pi 60 \text{ rad. sec.}^{-1})t + \frac{\pi}{2} \text{ rad.}\right]}{(2\pi 60 \text{ rad. sec.}^{-1})(200 \times 10^{-6} \text{ farad})} \\ &= \frac{10^4 \text{ coulomb-volts}}{8\pi \text{ coulomb}} \cos\left[(2\pi 60 \text{ rad. sec.}^{-1})t + \frac{\pi}{2} \text{ rad.}\right] \\ &= 398 \text{ volts} \cos\left[(2\pi 60 \text{ rad. sec.}^{-1})t + \frac{\pi}{2} \text{ rad.}\right]. \end{aligned}$$

(*c*) The maximum value of the back e.m.f. in the entire circuit is given by the vector sum of E'_{0R} and E'_{0C}. E'_{0R} and E'_{0C} are plotted in Fig. 247 in a vector diagram of the type shown in Fig. 238. E'_0 is the vector sum of E'_{0R} and E'_{0C} so that it has the value

$$E'_0 = [E'^2_{0R} + E'^2_{0C}]^{\frac{1}{2}} = [(450 \text{ volts})^2 + (398 \text{ volts})^2]^{\frac{1}{2}} = 601 \text{ volts}.$$

From Fig. 247, it may be seen that E'_0 leads I_0 by an angle $\pi - \varphi$ where

$$\begin{aligned} \varphi &= \arctan \frac{E'_{0C}}{E'_{0R}} = \arctan \frac{I_0/\omega C}{I_0 R} = \arctan \frac{1}{\omega C R} \\ &= \arctan \frac{1}{(2\pi 60 \text{ rad. sec.}^{-1})(200 \times 10^{-6} \text{ farad})(15 \text{ ohms})} \\ &= \arctan \frac{1 \text{ volt}}{377 \times 3 \times 10^{-3} \text{ sec.}^{-1} \text{ coulomb-ohm}} \\ &= \arctan 0.884 = 41°29' = 0.724 \text{ rad.} \end{aligned}$$

E' may therefore be written

$$\begin{aligned} E' &= 601 \text{ volts} \cos\,[(2\pi 60 \text{ rad. sec.}^{-1})t + \pi \text{ rad.} - 0.724 \text{ rad.}] \\ &= -601 \text{ volts} \cos\,[(2\pi 60 \text{ rad. sec.}^{-1})t - 0.724 \text{ rad.}] \end{aligned}$$

This is the back e.m.f. in the entire circuit. The applied e.m.fs. corresponding to E'_R, E'_C, and E' may be written thus, Fig. 247,

$$E_R = -E'_R = 450 \text{ volts} \cos (2\pi 60 \text{ rad. sec.}^{-1})t$$

$$E_C = -E'_C = 398 \text{ volts} \cos \left[(2\pi 60 \text{ rad. sec.}^{-1})t - \frac{\pi}{2} \right]$$

$$E = -E' = 601 \text{ volts} \cos [(2\pi 60 \text{ rad. sec.}^{-1})t - 0.724 \text{ rad.}].$$

The current in the entire circuit therefore leads the applied e.m.f. by an angle $\varphi = 0.724$ rad.

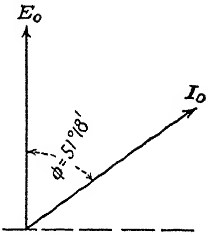


FIG. 248.—Vector diagram of applied e.m.f. and current. Illust. Prob. 3.

3. A 5-ohm resistance, a 40-millihenry inductance, and a 300-microfarad capacity are all connected in series. If an e.m.f. of 160 volts maximum value with a frequency of 60 cycles per second is applied to this circuit, find the current in the circuit and the power factor.

(*a*) The e.m.f. applied to the entire circuit is

$$E = 160 \text{ volts} \cos [(2\pi \text{ rad.})(60 \text{ sec.}^{-1})t]$$

The current in the circuit is, Eq. (27),

$$I = \frac{E_0}{Z} \cos (\omega t - \varphi)$$

where $E_0 = 160$ volts, the maximum e.m.f., $\omega = 2\pi 60$ the angular velocity in radians per second, Z is the impedance of the circuit. $L\omega$ is the inductive reactance, $1/C\omega$ the capacitative reactance, and $L\omega - \dfrac{1}{C\omega}$ is the total reactance. From Eq. (29), page 509

$$Z = \left[R^2 + \left(L\omega - \frac{1}{C\omega} \right)^2 \right]^{\frac{1}{2}}$$

$$= \left\{ (5 \text{ ohm})^2 + \left[(40 \times 10^{-3} \text{ henry}) (2\pi 60 \text{ rad. sec.}^{-1}) - \frac{1}{(2\pi 60 \text{ rad. sec.}^{-1})(300 \times 10^{-6} \text{ farad})} \right]^2 \right\}^{\frac{1}{2}}$$

$$= \left\{ 25 \text{ ohm}^2 + \left[15.08 \frac{\text{volt-sec.}}{\text{amp. sec.}} - 8.84 \frac{\text{volt-sec.}}{\text{coulomb}} \right]^2 \right\}^{\frac{1}{2}}$$

$$= [25 \text{ ohm}^2 + 38.9 \text{ ohm}^2]^{\frac{1}{2}} = (63.9)^{\frac{1}{2}} \text{ ohm} = 7.99 \text{ ohms}.$$

The angle φ, the angle by which the current lags behind the applied e.m.f. in phase, is given by

$$\varphi = \arctan \left(\frac{L\omega - \dfrac{1}{C\omega}}{R} \right) = \arctan \frac{15.08 \text{ ohms} - 8.84 \text{ ohms}}{5 \text{ ohms}}$$

$$= \arctan \frac{6.24}{5} = \arctan 1.248 = 51°18' = \frac{51.3° \times \pi \text{ rad.}}{180°} = 0.895 \text{ rad.}$$

Therefore,

$$I = \frac{160 \text{ volts}}{7.99 \text{ ohms}} \cos (2\pi\ 60 \text{ rad. sec.}^{-1}\ t - 0.895 \text{ rad.})$$
$$= 20 \text{ amp. } \cos (2\pi\ 60 \text{ rad. sec.}^{-1}\ t - 0.895 \text{ rad.}).$$

Figure 248 shows the current lagging behind the applied e.m.f. by an angle of 51°18′. The power factor is $\cos \varphi = \cos 51°18' = 0.625$.

4. A transformer which is 100 per cent efficient carries a 10-kw. load with a power factor of 0.866. The ratio of the number of turns on the secondary to the number on the primary is 100:1, and the transformer is connected

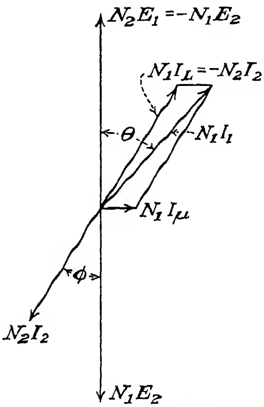


FIG. 249.—Vector diagram of 100 per cent efficient transformer, essentially the same as Fig. 243 of text. Illust. Prob. 4.

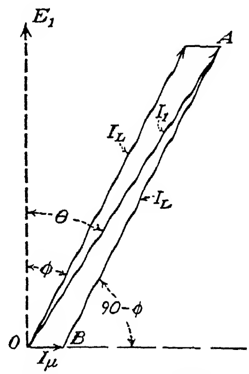


FIG. 250.—Vector diagram of currents in primary of transformer enlarged from part of Fig. 249. Illust. Prob. 4.

to a 230-volt 60-cycles per second line. The current in the primary is 52.5 amp. Find (*a*) the secondary e.m.f., (*b*) the secondary current, (*c*) the magnetizing current, and (*d*) the power factor of the primary current.

The vector diagram of a 100 per cent efficient transformer is given in Fig. 243. The part necessary for the solution of this problem is shown in Fig. 249. Each vector in Fig. 249 is multiplied by the appropriate number of turns, so that vectors representing e.m.fs. or currents in the secondary will have the same length as those representing the corresponding quantities in the primary.

In this problem, we are given

$E_1 = 230$ volts = the primary e.m.f.,

$I_1 = 52.5$ amp. = the total primary current,

$N_2/N_1 = 100$ = the ratio of the number of turns in the secondary coil to the number in the primary coil,

$\omega = 2\pi f = 2\pi \times 60$ rad. sec.$^{-1}$ = 377 rad./sec. = the angular velocity with which the vectors representing the current and differences of potential rotate about a common origin,

P_2 = 10,000 watts = the power of the load, and

$\cos\varphi = 0.866$ = the power factor of the load. From $\cos\varphi$, we find $\sin\varphi = 0.500$.

(*a*) The secondary e.m.f. E_2 is found from Eq. (37).

$$E_2 = \frac{N_2E_1}{N_1} = 100 \times 230 \text{ volts} = 23{,}000 \text{ volts}.$$

(*b*) The current in the secondary circuit is found from the power in an a-c circuit given by Eq. (34), as $P_2 = I_2E_2\cos\varphi$ or 10,000 watts = I_2 23,000 volts × 0.866 whence

$$I_2 = \frac{10{,}000 \text{ watts}}{23{,}000 \text{ volts } 0.866} = 0.502 \text{ amp.},$$

the current in the secondary circuit. From Par. 31, the load component of the primary current I_2, Fig. 249, is given by $I_2 = I_L\frac{N_1}{N_2}$ from which I_L = 0.502 amp. 100 = 50.2 amp.

(*c*) We now calculate the magnetizing current I_μ from the vector diagram. The current diagram enlarged from Fig. 249 is shown in Fig. 250. From the law of cosines,

$$I_1^2 = I_\mu{}^2 + I_L^2 + 2I_\mu I_L\cos(90 - \varphi)$$
$$I_\mu{}^2 + (2I_L\sin\varphi)I_\mu + I_L^2 - I_1^2 = 0$$
$$I_\mu{}^2 + 2 \times 50.2 \text{ amp.} \times 0.500I_\mu + (50.2 \text{ amp.})^2 - (52.5 \text{ amp.})^2 = 0$$
$$I_\mu = \tfrac{1}{2}(-50.2 \text{ amp.}) \pm \tfrac{1}{2}\{(50.2 \text{ amp.})^2 - 4[(50.2 \text{ amp.})^2 - (52.5 \text{ amp.})^2]\}^{\frac{1}{2}}$$
$$= (-25.1 + \tfrac{1}{2}\sqrt{3{,}465}) \text{ amp.} = (-25.1 + \tfrac{1}{2}58.9) \text{ amp.} = 4.4 \text{ amp.}$$

(*d*) The power factor of the primary current is $\cos\theta$ where θ is the angle between I_1 and E_1 in Fig. 250.

$$\cos\theta = \frac{I_L\cos\varphi}{I_1} = \frac{50.2 \text{ amp. } 0.866}{52.5 \text{ amp.}} = 0.828.$$

PROBLEMS

1. A car moves on a level track under the influence of a periodic applied force of 5 newtons amplitude. Friction is the only force opposing the motion of the car, and the frictional resistance constant is 2 kg./sec. Find the velocity of the car when the force is a maximum and zero.

2. A 50-kg. car moves on a level track under the influence of a periodic applied force of 15 newtons amplitude and 2 sec. period. There is neither friction nor restoring force. Find the velocity when the force is a maximum, zero, and 10 newtons.

3. A car moves on a level track under the influence of a periodic applied force of 4 newtons amplitude and a frequency of 20 vibrations per second; and an elastic restoring force with a constant of 50 newtons/m. Find the velocity when the force is a maximum, zero, and 0.01 sec. after the force has its maximum value.

4. A 60-cycle 160-volt peak e.m.f. is applied to a resistance of 10 ohms. Find the rate of flow of charge through the resistance 0.003 sec. after the e.m.f. passes through zero as it is increasing.

5. An inductance of 20 millihenries is connected to a 160-volt peak 60-cycle line. Find the phase of the e.m.f. when the current is -10 amp. and decreasing. What is the value of the e.m.f. at the same instant?

6. A 60-cycle 160-volt peak e.m.f. is applied to a condenser of 50 microfarads capacity. What will be the value of the e.m.f. when the current is 1 amp. and increasing?

7. Given a resistance of 3 ohms, an inductance of 60 millihenries, and a capacity of 50 microfarads connected in series, what is (*a*) the inductive reactance, (*b*) the capacitative reactance, and (*c*) the total reactance? If a current $10 \cos 360t$ amp. flows in these impedances, what is the value of the difference of potential across (*d*) the resistance, (*e*) the inductance, and (*f*) the capacity? Find (*g*) the total e.m.f., and draw a vector diagram representing the current and all the e.m.fs.

8. What is the impedance of a 60-cycle circuit containing 60 ohms resistance, 50 millihenries inductance, and 80 microfarads capacity in series? What is the phase of the applied e.m.f. relative to the current in this circuit?

9. What is the value of the capacity which in series with 100 ohms resistance and 20 millihenries inductance will produce a 60-cycle current 30° ahead of the applied e.m.f. in phase?

10. The total reactance of a series circuit containing resistance, inductance, and capacity is 30 ohms. Find the peak value of the current which lags 30° behind the peak e.m.f. of 320 volts.

11. What is the effective value of a peak voltage of 160 volts? If an alternating current produces 9.56 cal./sec. in a 10-ohm resistance, what is its effective value?

12. A solenoid has a resistance of 10 ohms and an inductive reactance of 90 ohms. What is the power factor, and how much power does it consume from a line with an effective difference of potential of 115 volts?

13. What power is saved by transmitting 200 kw. at 2,300 volts instead of 115 if the line resistance is $\frac{1}{2}$ ohm and the power factor 0.8?

14. A transformer in which there are no losses with a primary inductance of 0.5 henry is connected with the secondary on open circuit to 115-volt effective 60-cycle mains. What is the value of (*a*) the e.m.f. applied to the primary, (*b*) the back e.m.f. induced in the primary, (*c*) the phase angle between these two e.m.fs., (*d*) the magnetizing current, and (*e*) the phase angle between the magnetizing current and the applied e.m.f.?

15. A transformer which is 100 per cent efficient has 1,000 primary turns and 5,000 secondary turns. The primary is connected to a 230-volt effective 25-cycles/sec. line. The secondary current is 10 amp. with a lagging power factor of 0.8. The magnetizing current is 1 amp. Find (*a*) the secondary

e.m.f., (*b*) the impedance of the load, (*c*) the load component of the primary current, (*d*) the primary current, and (*e*) the rate of change of flux in the core.

16. A transformer which is 100 per cent efficient carries a load of 200 ohms resistance and 0.5 henry inductance. The ratio of secondary to primary turns is 10:1 and the primary, the inductance of which is 1 henry, is connected to a 115-volt effective 60-cycle line. Find (*a*) the power factor of the load, (*b*) the magnetizing current, (*c*) the secondary current, (*d*) the primary current, (*e*) the angle of lag of the primary current, (*f*) the mutual inductance, and (*g*) the inductance of the secondary coil.

17. Compute the power consumed by the primary of the transformer of Prob. 16, and show that it is equal to the power delivered to the load.

CHAPTER XXV

WAVES AND WAVE MOTION

1. Although we discussed several different forms of energy in the early chapters of this book and explained how one type of energy might be converted into another, we paid little attention to methods of transmitting energy. One of the most important modern means of transmitting energy is by alternating currents, a topic treated in the last chapter, but this is a method only recently developed and one that does not occur in nature. The principal means by which energy is transmitted in nature is by trains of waves. This is the way in which the sun's energy reaches us and is therefore of fundamental importance since the sun is the ultimate source of most of the energy that we use in daily life. At present, only small amounts of energy are transmitted by man-controlled trains of waves, but since this type of energy transmission includes sound and radio communication it is of great practical importance. We shall therefore proceed to a general treatment of wave motion, applying the knowledge of mechanical and electrical oscillations that we have presented in the last few chapters.

2. The most familiar waves are those on the surface of a body of water. When such waves are produced by the action of winds over large areas, it is difficult to follow in detail the process of formation of the waves or to localize the place of origin and place of ultimate dissipation of the energy. A much simpler and smaller scale example of wave motion is found in the waves spreading out from the point where a stone is dropped into a still pond. On striking the surface of the water, the stone is immediately retarded by the inertia and viscous resistance of the water. Part of the stone's kinetic energy is changed directly into heat by the viscous resistance, but much of it goes into pushing aside the water under the stone. This water piles up in a circle around the stone somewhat in the fashion shown in cross section in Fig. 251. Gravitation immediately furnishes a restoring force which

tends to bring the water back to its level equilibrium position. Because of inertia, the water goes past its equilibrium position, executing a damped oscillation about this position. This water right at the place where the stone fell is connected with all the other water in the pond by intermolecular forces. On the surface,

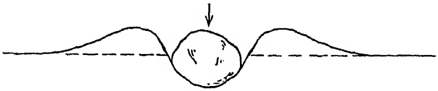


FIG. 251.—Cross-sectional diagram showing the effect of a stone striking a water surface.

these forces manifest themselves in the surface tension, under the surface, in viscosity and cohesion. Evidently the motion of one part of the water will affect other parts through the medium of these forces. Consequently, as everyone knows, the disturb-

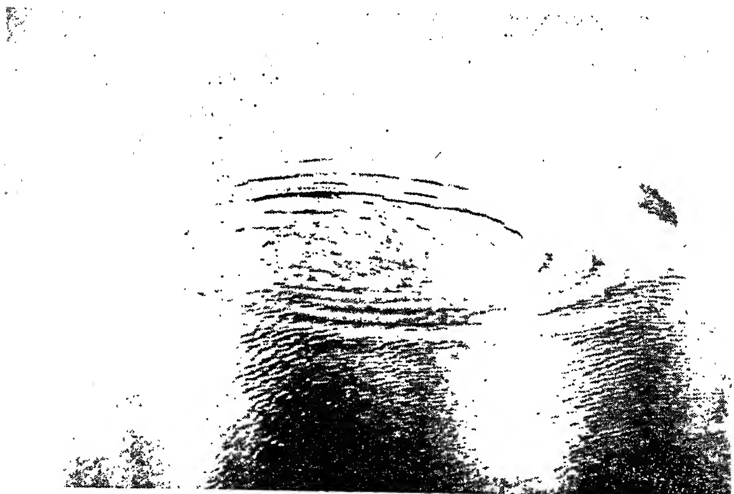


FIG. 252.—Photograph of waves spreading out from the place where a stone has struck water.

ance from the stone spreads out over the surface of the water in a series of circular waves. Casual observation shows that there are usually about half a dozen noticeable waves in the train of waves resulting from a single splash. This wave train will impart motion to anything floating on the water but such motion is one

of oscillation which gives no resultant displacement to the floating object. That the wave train transmits energy is apparent from the vigor with which the waves splash on the shore.

Waves on a Rope

3. As another simple example of wave motion, consider the waves that can be sent along a slack rope. Suppose that one end of the rope is fastened to a post and the other held in the hand. If the free end is suddenly jerked sideways, the distortion produced travels down the rope and is reflected back from the fixed end. If the free end is moved back and forth several times,

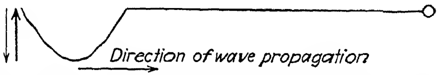


FIG. 253.—A wave impulse starting down a rope.

a series of distortions or waves travels down the rope. The same effect can be more clearly observed in a long spiral spring suspended from the ceiling.

Transverse and Longitudinal Waves

4. In both water waves and waves in a rope, the principal motion is sideways, perpendicular to the direction in which the waves travel. Such waves are called transverse waves. But there is another kind of wave in which there is an oscillatory motion parallel to the direction of propagation of the waves. Such a wave is called a **longitudinal** wave. A loosely coiled spiral spring hung from the ceiling can transmit a longitudinal wave as well as a transverse one. If the lower end of such a spring is compressed and then released, a wave of longitudinal compression will travel along the spring to the ceiling, be reflected from the support at the top, and travel down again. This phenomenon can probably be understood better in terms of another experiment. Suppose that a number of heavy balls are connected in a line by small coil springs as shown in Fig. 255. Suppose that the end one is pushed to the right. As soon as it has moved a little, but not before, the spring is somewhat compressed and exerts a force on the second ball which then begins to move. This in turn compresses the second spring and moves the third

ball. In this way, a wave of compression travels down the line of balls. Because of the inertia of the balls and the compressibility of the springs, it takes a finite time for the effect of the motion of the first ball to be transmitted down the line. The lighter the balls and the stiffer the springs, the faster this wave of compression will travel. Suppose now that the first ball is not merely moved once to the right but is kept oscillating first to the right and then to the left. The first spring and subsequently all the others will then be alternately compressed and stretched, and a train of waves of alternate compressions and extensions will travel down the line of balls. The continuous spring that we first described behaves in essentially the same way except that every part of the spring contributes both to the inertia and to the elasticity of the system. The system of heavy balls and light connecting springs might be used equally well for the illustrations of a transverse wave or can be given a complex motion transmitting a wave that is partly transverse and partly longitudinal. Water waves have this complex character.

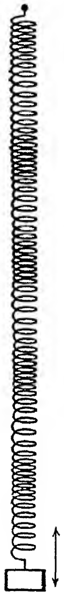


FIG. 254.—Longitudinal compressional waves on a vertical spiral spring.

5. It is clear that in a continuous medium like water where there is no elastic resistance to shear there can be no transverse waves. (On the surface of water the force of gravity acts as a restoring force, taking the place of elasticity. We are speaking now about waves in the body of a uniform medium.) But in a solid which by definition has resistance to shearing stresses, a transverse wave can be transmitted.

6. Longitudinal waves, on the other hand, can be transmitted by any medium that offers elastic resistance to compression. In a gas, for example, which cannot transmit a transverse wave, a wave of compression or rarefaction can be transmitted. Sound consists of waves of this sort.

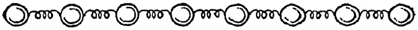


FIG. 255.—Heavy balls connected by springs. An arrangement capable of transmitting either longitudinal or transverse waves.

Characterization of a Wave

7. The first type of wave that we shall attempt to describe will be a transverse wave, or rather train of waves, traveling down a string so long that there is no reflected wave. In Fig. 256, we given an instantaneous picture of the string. Think of this drawing first merely as a snapshot of the string with a straight line added to indicate the position of the string when there was no wave present. Then the perpendicular distance from this line to any point of the string gives the displacement of that point from its original undisturbed position. For convenience, we can call this displacement s. The maximum displacement that ever

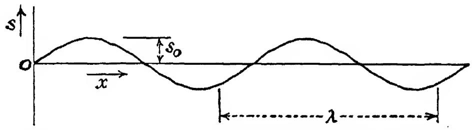


FIG. 256.—Instantaneous picture of a wave on a string, or graph of s vs. distance where s is any physical quantity propagated as a wave.

occurs at any point of the string is called the amplitude of the wave. Any two points on the string that have the same displacement and are moving in the same direction are said to have the same phase, or to be in phase. The distance between any two such points that are adjacent is called the wave length and designated by the Greek letter λ (lambda). Assume that we are dealing with a sine wave. This means that the displacement s of any point along the wave that is at a distance x from a point of zero displacement is given by an equation of the form

$$= s_0 \sin \frac{2\pi x}{} \tag{1}$$

where s_0 is the amplitude. This may seem a very special assumption, but it was shown by Fourier that any periodic disturbance could be considered the resultant of a number of sine waves of different wave lengths and amplitudes. We shall consider only cases where a single sine wave is sufficient to describe the motion completely.

8. We have suggested that Fig. 256 be considered first as a direct picture of the vibrating string. Now we wish to point out that it can equally well be considered as a graph of the displace-

ment s as a function of the distance x from some fixed point, say O, of the string. From this point of view, the disturbance which is propagated as a wave and plotted in Fig. 256 need not be a mechanical transverse displacement. It may be a pressure, a temperature, an electric field strength, or anything else that is varying from place to place, according to a sine law as given in Eq. (1).

9. The graph in Fig. 256 shows the way in which the disturbance varies with distance along the string at a given instant of time. We next investigate the disturbance at a particular point on the string as the time changes. If the wave is traveling at a constant

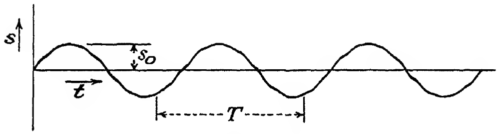


FIG. 257.—Graph of a wave disturbance s at a given point in space, as a function of time.

speed V, the disturbance at any particular point is one which was at a distance $x = Vt$ away t sec. earlier. During an interval of time λ/V, the disturbance at a particular point must go through all the values it had in the range of distances zero to λ away at the beginning of the time interval. Since these values include all possible ones as given by Eq. (1), the time variation at the particular point corresponds to the space variation given by Eq. (1) and must be expressed in a similar form. Thus the time variation for a particular point is

$$= s_0 \sin \frac{2\pi t}{\lambda/V} \qquad s = s_0 \sin \frac{2\pi t}{T} \qquad (2)$$

where T is called the period of the wave. This time variation given by Eq. (2) is plotted in Fig. 257 where the scale of the sine wave has been changed deliberately in order to emphasize the difference between this graph and that of Fig. 256. In Fig. 257, the period T is the interval along the time axis between any point and the next one on the curve which has the same phase.

10. In Fig. 256 and Eq. (1), we expressed the way in which the disturbance s varies from point to point along the direction of propagation of the wave at any given instant. In Fig. 257 and

Eq. (2), we expressed the way in which the disturbance at any given point varies as time passes. It is useful to combine these two variations in a single relation which expresses the variation of the disturbance both in time and in space. Let us begin with Fig. 256 and say that it indicates the condition at the time $t = 0$. If we are dealing with a continuous train of waves of constant amplitude, moving with a velocity V, the condition at any later time t' will be obtained by shifting the entire curve a distance Vt' to the right. If $Vt' = \lambda$, one wave length, then the shifted curve will be identical with the original one. If $Vt' = \lambda/2$, the shifted curve will look like the original curve inverted, and so on. But we know from Fig. 257 and Eq. (2) that the original condition of the wave is repeated when the time is increased by the period T. Therefore if Vt' is equal to λ, we must have $t' = T$ and therefore $VT = \lambda$. Suppose, as we have tacitly assumed, that the wave is traveling in the positive direction of x. If $t - t_0$ is the time it takes for the disturbance to travel from the point $x = 0$ to the point $x = x$, then the disturbance at x at the time t will be the same as the disturbance at $x = 0$ at the time t_0. Thus if $s = s_0 \sin 2\pi t_0/T$ is the disturbance at $x = 0$ when $t = t_0$, the disturbance at $x = x$ at the time t will be the same but may be written $s = s_0 \sin \dfrac{2\pi[t - (t - t_0)]}{T}$. But $t - t_0 = x/V$ so that by substitution we get

$$s = s_0 \sin \frac{2\pi}{T}\left(t - \frac{x}{V}\right).$$

Since $VT = \lambda$, this can be written

$$s = s_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right), \tag{3}$$

an equation that gives the disturbance at any time t at any point x along the direction of propagation of the wave in terms of the amplitude s_0, the period T, and the wave length λ. We have already defined wave length and period. We shall now define the frequency of a wave as the number of complete oscillations that occur at any given point in 1 sec. or the number of wave crests that pass any given point in 1 sec. The frequency is usually denoted by f or by the Greek letter ν (nu) and is equal to $V/\lambda = 1/T$.

Transmission of Energy by Waves

11. It is no news to anyone who has watched the waves of the sea beating on a coast to hear that waves transmit energy. All light energy and all the heat energy from the sun are transmitted by electromagnetic waves. For any given type of wave, the amount of energy transmitted is proportional to the square of the amplitude of the wave.

Sound Waves

12. As we have already mentioned, sound waves are compressional waves. All noises and most musical sounds are combinations of waves of many different frequencies. In musical sounds, there are definite numerical relationships between the frequencies occurring. The pitch of the sound is ordinarily determined by the lowest frequency present, but the quality depends on the relative amplitudes of all the frequencies present. A tuning fork gives out a wave that has only one frequency in it.

Doppler Effect

13. If the source of a wave and its receiver are moving relative to each other, the number of waves reaching the receiver per second is altered. If the source and receiver are approaching, this number is increased; if they are separating, it is decreased. Examples of this effect are often encountered when two trains or two cars pass each other at high speeds. If the whistle or automobile horn is sounding at the time, a sharp drop in pitch can be noticed when the relative motion changes suddenly from approach to recession. This effect is called the Doppler effect and can be treated exactly in a perfectly straightforward way.

14. If the ear stands still and listens to waves of a certain frequency n, then n waves strike the ear in 1 sec. Now if the ear moves toward the source, an additional number of waves will strike it per second which will be equal to the distance the ear moves toward the source in 1 sec., *i.e.*, the velocity of the ear v_e times the number of waves in unit distance. The number of waves in 1 m. is $1/\lambda$, where λ is the wave length in meters, so that the number of waves in v_e m. is v_e/λ, which is therefore the increase in frequency due to the motion of the ear. The final frequency is thus the original plus the increase or

$$n' = n + \frac{v_e}{\lambda}. \tag{4}$$

Now in any wave motion $\lambda = V/n$, which when substituted in Eq. (4) gives

$$n' = n + \frac{nv_e}{V} = n\left(\frac{V + v_e}{V}\right). \tag{5}$$

If the ear recedes from the source, a similar argument shows that

$$n' = n\left(\frac{V - V_e}{V}\right). \tag{6}$$

15. If the source stands still, the number of waves it emits in 1 sec. n will occupy the space between the source and a point at a distance from the source equal to the distance V which the waves

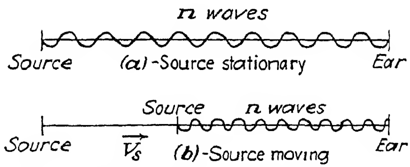


FIG. 258.—The Doppler effect for sound. The motion of the source increases the number of waves per second reaching the ear.

travel in 1 sec. The length of each wave is equal to the distance V which they occupy divided by the number of waves n, or $\lambda = V/n$ as usual. Now if the source moves toward the ear in the direction in which the waves are traveling with a velocity v_s, in 1 sec. it will have caught up a distance v_s on the first wave. Therefore the n waves which were emitted during that second will now lie between a point at a distance V and a point at a distance v_s from the original position of the source as shown in Fig. 258. Therefore the n waves emitted during 1 sec. have been compressed into a distance $V - v_s$, and the wave length is now

$$\lambda'' = \frac{V - v_s}{n} = \frac{V}{n''}, \tag{7}$$

so that

$$n'' = n\left(\frac{V}{V - v_s}\right). \tag{8}$$

Again, if the source recedes from the ear, a similar argument shows that

$$n'' = n\left(\frac{V}{V + v_s}\right). \tag{9}$$

16. Finally, if both the ear and the source move toward each other, the final frequency will be obtained by using in Eq. (5), as the frequency n of the waves in the neighborhood of the ear, the frequency n'' from Eq. (8). This gives for the frequency heard, when both ear and source approach each other,

$$n' = n\left(\frac{V + v_e}{V - v_s}\right). \tag{10}$$

In the general case of either approach or recession of either source or ear, this becomes

$$n' = n\left(\frac{V \pm v_e}{V \mp v_s}\right)\begin{cases}\text{upper sign approach}\\ \text{lower sign recession.}\end{cases} \tag{11}$$

Reflected Waves. Change of Phase

17. We have seen that waves on a rope or spring are reflected at the ends. Similarly a water wave striking a wall is reflected. In general, any sharp change in the medium through which a wave is traveling will produce total or partial reflection. We shall now discuss an example of such phenomena in some detail.

18. Consider a spiral spring hung from the ceiling (Fig. 254). Suppose a single pulse of compression followed by extension is traveling down the spring. Each turn of the spiral moves downward, giving momentum to the next lower turn until finally the lowest turn moves downward. Since this end turn is hanging free, there is no force from below to resist its motion and it will continue to move until the restoring elastic forces from above stop it and then reverse its motion. This will not occur until the lower end of the spring has been extended. The return toward equilibrium from this extended condition will send a wave back up the spring. So much for reflection at the free end. Now suppose that the compressional pulse is traveling up the spring toward the fixed support at the top. As the upward momentum reaches the topmost turn, it produces no motion because the upper end of the spring is fixed and cannot move. The fixed

support reacts against the upward force of the spring with an equal and opposite force. Essentially we have a perfectly elastic collision with a body of infinite mass. The momentum of the upper turns of the spring is reversed without this part of the spring being extended in contrast to the reflection at the free end where the reflection is the result of unopposed extension. The reflected wave from the free end acts exactly as if it were the original wave traveling in the opposite direction, but the reflected wave at the fixed end starts back with a compression before ever having gone through an extension. In terms of phase, the wave reflected at the free end is in phase with the incident wave, that reflected at the fixed end is 180° ahead of the incident wave.

19. Similar considerations apply equally well to transverse waves. At the fixed end, the spring cannot move sideways and the transverse momentum brought to the end of the spring by the wave is reversed in direction by the reaction of the support. This reversal can easily be seen in a horizontal rope or spring fastened to a wall at one end and held in the hand at the other. An upward distortion of the rope will be reflected back as a downward distortion, and vice versa. Another way of looking at this reversal may be helpful. Think of the rope as passing through a small hole in the wall so that it cannot have any appreciable motion but is not actually fastened. Then imagine the reflected wave replaced by a wave coming along the rope from the other side of the hole. This wave is to combine with the incident wave in such a way as to give the same effect as the reflection. The first condition to be satisfied is that no resultant motion occurs at the hole. The new wave must therefore be just opposite in shape to the incident distortion, and they must reach the hole simultaneously so that their effects cancel each other. The new distortion coming from behind the hole will then continue toward the hand as a distortion exactly opposite in shape to the original one.

20. In gravitational water waves reflected at a vertical surface, the transverse component moves up and down along the surface without encountering resistance and is reflected without change of phase. In general, the question of the phase of the wave reflected from the surface between two mediums depends on the characteristics of the two mediums pertinent to the transmission of the particular type of wave. Such properties may include

the density, elasticity, dielectric constant, resistivity, and so on.

21. Mathematically, if the incident wave is represented by Eq. (3), the reflected waves are represented by

$$s' = s_0 \sin 2\pi\left(\frac{T}{t} + \frac{x}{\lambda}\right) \text{ for no change of phase} \tag{12}$$

and

$$\begin{aligned} s' &= s_0 \sin\left[2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right) + \pi\right] \\ &= -s_0 \sin 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right) \text{ for a change of phase of } 180° \end{aligned} \tag{13}$$

where the positive sign of the x/λ term indicates that the wave is traveling in the negative direction along the x-axis.

Superposition of Waves

22. If two stones are thrown simultaneously into a pond, the wave systems from the two splashes will generally spread out toward each other. When they meet, the motion of the water is complicated, but they soon pass through each other and appear on the other side unmodified. This is a general characteristic of waves. The presence of one wave train in a medium does not prevent or modify the passage of another. At any particular point at any instant, the resultant effect is simply the sum of the two separate effects. The resultant effect may be greater or less than either of the separate effects, depending on the phase relations. Such effects are said to be caused by interference.

Standing Waves

23. One of the simplest examples of interference occurs when waves of the same frequency are traveling in opposite directions. Such a condition can be easily obtained if a wave is reflected back along the direction of propagation. Suppose, for example, that the end of a light chain is moved up and down with simple harmonic motion. A wave travels down the chain and is reflected at the other end which is fixed. Suppose that the length of the wave is ust equal to the length of the chain. As we have seen the wave reflected at the fixed end is changed in phase by 180°, or half a wave length. The resultant effects can be understood by studying the successive stages of the waves as shown in Fig. 259.

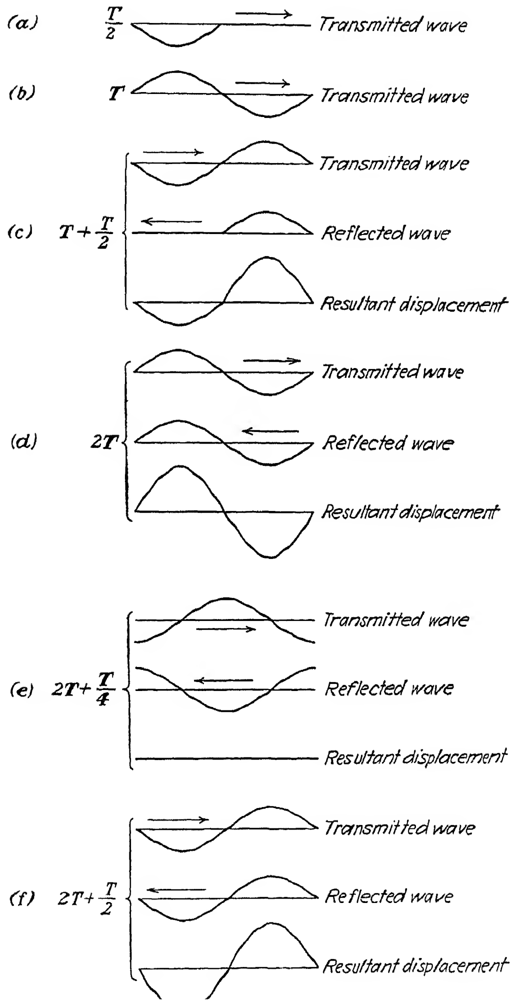


FIG. 259.—Successive stages in the formation of standing waves on a chain. (a), (b), etc., show the component effects of the transmitted and reflected waves and the resultant effect at times $T/2$, $2T/2$, etc. after the initial wave starts down the chain from the left. The wave length is taken equal to the length of the chain.

It is seen that there is one place in the center of the chain where there is never any motion. The two waves always arrive here 180° out of phase, or as we say, in opposite phase so that their resultant effect is zero. Such a point is called a node. On the other hand, there are two places one-fourth and three-fourths of the way down the chain where the motion is a maximum. These points are called antinodes. The resultant motion of the chain is said to be a standing wave. If the length of the wave sent down the chain is gradually diminished, the standing waves disappear and the resultant motion of the chain becomes chaotic until the wave length $\lambda = \frac{2}{3}L$ where L is the length of the chain. At this point, standing waves are set up again. Figure 260 illustrates this case and shows that there are two nodes and three antinodes. Further consideration or experimental test shows that standing waves appear whenever the length of the chain is equal to an integral number of half wave lengths, *i.e.*, if $L = n(\lambda/2)$. Furthermore, the number of nodes in addition to those at the ends will be given by $n - 1$ and of antinodes by n. We have discussed standing waves on a chain in detail, but the same considerations apply to any kind of standing wave.

24. In general, standing waves are produced whenever two trains of waves of the same velocity of propagation, wave length and amplitude are traveling in opposite directions in the same region. By choosing any arbitrary origin of coordinates and calling x the distance from this origin along the line of propagation of the waves, the two waves are represented by

$$s = s_0 \sin\left[2\pi\left(\frac{T}{t} - \frac{x}{\lambda}\right) + \delta\right] \tag{14}$$

$$s' = s_0 \sin\left[2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right) + \delta'\right] \tag{15}$$

where the first equation represents the wave traveling in the positive direction of x and the second, with the positive sign, that traveling in the negative direction of x. The angles δ and δ' are constants depending on the choice of the origins of coordinates and of time. By a process similar to that of Chap. XXIII, Pars. 21–23, it is possible to get these equations in a form that shows the presence of nodes and antinodes.

25. We shall simplify the problem by assuming that one of the two wave trains is caused by the reflection of the other and by

choosing the point of reflection as the origin. Also we shall start measuring the time from an instant when the incident wave is

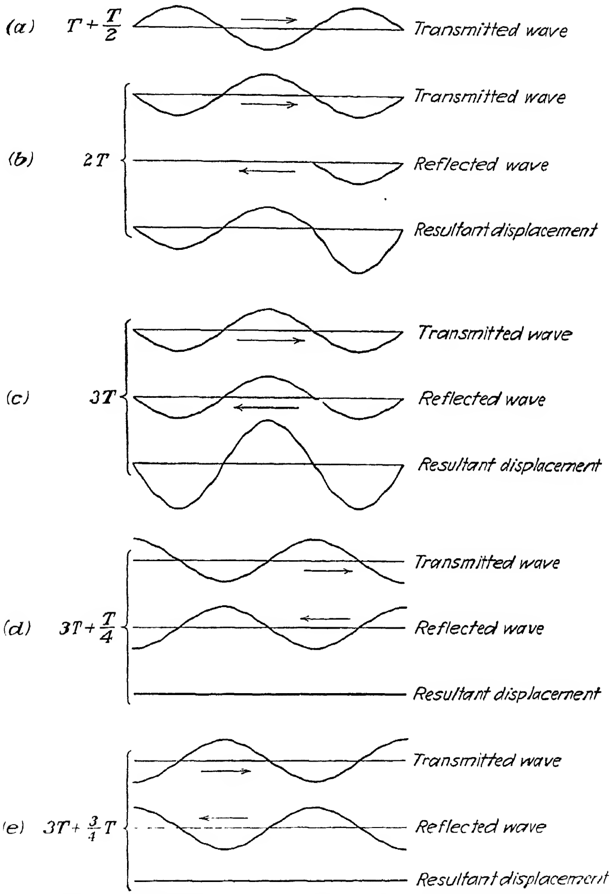


FIG. 260.—Similar to Fig. 259, but here the wave length is taken as two-thirds the length of the chain.

just beginning a cycle at the origin. Further, we shall call the positive direction of x the direction away from the reflection point so that the incident wave is traveling in the negative direction

and the reflected wave in the positive direction. The equation of the incident wave is then

$$s = s_0 \sin 2\pi\left(\frac{t}{T} + \frac{x}{\lambda}\right). \tag{16}$$

For the reflected wave two cases must be distinguished, (*a*) reflection without change of phase and (*b*) reflection with a change of phase of 180° (see Pars. 17 to 19). The equations for the reflected waves in the two cases are

$$(a) \quad s' = s_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) \text{ (no phase change)} \tag{17}$$

$$(b) \quad s' = s_0 \sin\left[2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) + \pi\right] \text{ (phase change)} \tag{18}$$

$$= -s_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right).$$

To get the resultant effect, the appropriate one of these equations must be added to (16) for the incident wave. For case (*a*) by expanding and adding we get for s_R, the resultant disturbance at a distance x from the point of reflection at the time t,

$$s_R = s + s' = s_0\left[\sin\frac{2\pi t}{T}\cos\frac{2\pi x}{\lambda} + \cos\frac{2\pi t}{T}\sin\frac{2\pi x}{\lambda} + \sin\frac{2\pi t}{T}\cos\frac{2\pi x}{\lambda} - \cos\frac{2\pi t}{T}\sin\frac{2\pi x}{\lambda}\right] = 2s_0\sin\frac{2\pi t}{T}\cos\frac{2\pi x}{\lambda} = \left(2s_0\cos\frac{2\pi x}{\lambda}\right)\sin\frac{2\pi t}{T} \tag{19}$$

This is the equation of a standing wave and shows that the disturbance at any point x oscillates with the period T between the limits $+2s_0\frac{\cos 2\pi x}{\lambda}$ and $-2s_0\frac{\cos 2\pi x}{\lambda}$. If $2\pi x/\lambda = \pi/2,\ 3\pi/2 \cdots$, *i.e.*, if $x = \lambda/4, 3\lambda/4 \cdots (n+1)\lambda/2 \cdots$, $\cos 2\pi x/\lambda = 0$, there is no disturbance at any time and we have a node. If $2\pi x/\lambda = 0,\ \pi,\ 2\pi \cdots$, *i.e.*, if $x = 0,\ \lambda/2,\ \lambda \cdots n\lambda/2 \cdots$, $\cos\frac{2\pi x}{\lambda} = \pm 1$ and the amplitude is a maximum, corresponding to an antinode. Thus, in this case of reflection without change of phase, the point of reflection is an antinode.

26. For reflection with change of phase, Eq. (18) must be added to (16). The process is identical with that just carried out except for a change in the sign of the last two terms of the expanded expression for s_R. Because of this change, the resultant value of s_R is

$$s_R = 2s_0 \sin \frac{2\pi x}{\lambda} \cos \frac{2\pi t}{T} \tag{20}$$

which shows the nodes and antinodes interchanged with the corresponding positions in the previous case. Evidently the nodes

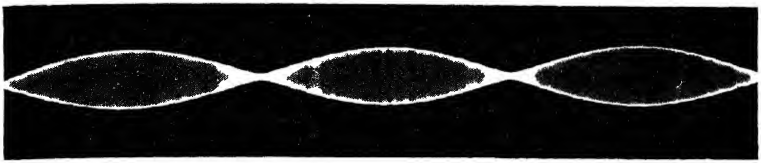


FIG. 261.—Photograph of standing waves on a chain. The distance between nodes was about 0.5 m.

are now at $x = 0,\ \lambda/2,\ \cdot\cdot\cdot\ ,\ n\lambda/2$ and the antinodes are at $x = \lambda/4,\ 3\lambda/4,\ \cdot\cdot\cdot\ ,\ (n+1)\lambda/2\ \cdot\cdot\cdot$. These results correspond to the standing waves set up on a chain fixed at the ends as shown in Fig. 261 and discussed in Par. 23.

Beats

27. Another type of simple interference phenomenon is illustrated by the resultant effect of two tuning forks F_1 and F_2 of slightly different periods. The sound from two such forks is pulsating. This is because there is a constant change in the relative phase of the waves coming from the two forks. When they are exactly in phase, there is a maximum of sound; when they are exactly 180° out of phase, there is a minimum. If the frequencies of the two forks are f and f' vibrations per second, the number of beats per second is $f - f'$.

Polarization of Transverse Waves

28. We have defined a transverse wave as one where the disturbance was perpendicular to the direction of propagation of the wave. This implies that the disturbance is a displacement or a force or a field strength, something that has a definite direction and can therefore be represented by a vector. But when we say

that the wave is transverse, we have defined only the plane in which this vector must lie, not its specific direction. If the direction of this vector remains constant, the wave is said to be polarized. Thus a water wave is polarized since the motion of any part of the water is in a vertical plane. A transverse wave along a string may be polarized in any plane or not polarized at all.

Electromagnetic Waves

29. In an a-c transmission line, the electrons that carry the current do not flow continuously from one end of the line to the other. They flow back and forth first in one direction and then in the other. If the frequency of alternation is high, the distance any one electron moves in one cycle, or period of alternation, may be very small. Thus the electrons move back and forth under the action of an alternating force somewhat as the balls in Fig. 255 move back and forth under the action of a longitudinal wave. The energy transmitted by an alternating current is carried by this oscillation of the electrons just as the energy of water waves is transmitted by the oscillation of the water without any change in the average position of the water. There is of course one great difference between the motion of the water and of the electrons. We can see the water move, but we cannot see the electrons. All we know about the electrons is that we apply an alternating e.m.f. at one point in a circuit and get an effect at some other point in the circuit. We interpret this in terms of an alternating electric field all along the wire and assume that the electrons in the wire respond to this field.

30. But we have seen that no wires are necessary for the transmission of electric and magnetic fields. One charge can act on another even if the space between them is entirely free of matter; a cathode-ray beam responds to the magnetic field of a neighboring current no matter how highly we evacuate the intervening space. If one plate of a condenser is charged, an electric field is set up all through the space around the plate and charges are induced in neighboring conductors. If the charge on the condenser plate is reversed, the direction of the field around it is reversed and the sign of the charges induced in the neighboring conductors is reversed. Similarly, if a current is flowing in a wire, a magnetic field is set up which reverses in direction when the current in the

wire reverses. We now wish to consider the question as to whether these reversals in the electric and magnetic fields follow the reversals of charge and of current *immediately* or whether there is a time lag corresponding to a finite speed of propagation of these effects through space. Certainly the time lag is very small, or we would have detected it already in some of the experiments we have performed. As far as we could see, the cathode-ray beam of the oscillograph followed the movements of a magnet instantaneously.

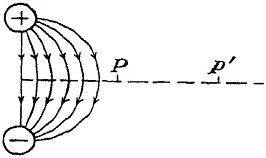


FIG. 262.—The electric field spreading out from two spheres that have just been charged. Only a few lines of force are shown.

31. Let us assume that the changes of electric and magnetic fields in space cannot keep up instantaneously with the changes in the charges and currents which give rise to them. To simplify our discussion, suppose we are dealing with the electric field of two charged spheres (see Fig. 90(*b*), page 230). Suppose that the spheres are originally uncharged, and therefore the electric field at any distant point P is zero. The spheres are then charged, but the field at P remains zero for a finite time before growing to the value corresponding to the charges put on the spheres. Call the time necessary to establish this field at P, t. Suppose that during this time of t sec. the charges on the condensers have reversed. It will again take t sec. for the field of these new charges to be felt at P, and during this second time interval the original field will have spread out to another point P' twice as far away from the spheres as P. If we continue to reverse the charge on the spheres, the field at P, P' and every other point in space will continue to reverse; but the field at P will always be opposite to that at the spheres; and the field at P' will always be opposite to that at P but the same as that at the spheres and so on. (Perhaps the argument is clearer if we specify that P and P' are along a line perpendicular to the line joining the spheres and through its center.) In other words, the electric field travels out from the spheres as a wave. If, then, electric and magnetic fields are propagated with finite speeds, oscillating currents and charges must send out waves of electric and magnetic field strength and we should be able to detect some of the properties of waves in the disturbances sent out by oscillating circuits. In

fact, this may be the easiest way to demonstrate that the speed of propagation of such disturbances is finite.

32. In Chap. XI, Par. 28, we defined an electric doublet, or dipole, as a system of two opposite charges whose separation was small compared with the distances to the points where their effects were to be considered. Similarly, if we are sufficiently

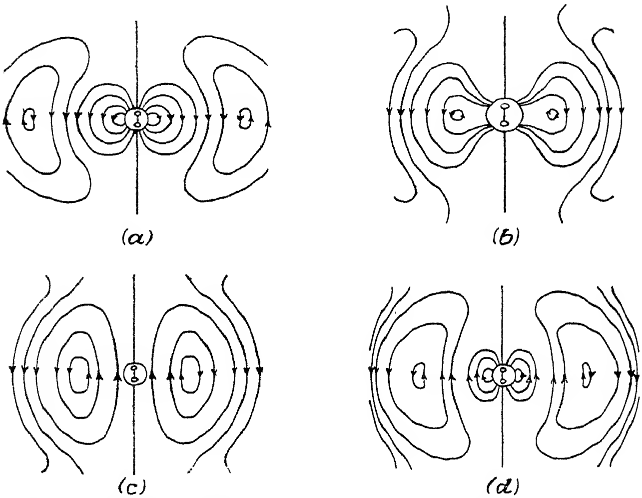


FIG. 263.—Successive stages of the electric field sent out by an oscillating electric dipole. The axis of the dipole is vertical. There is an accompanying magnetic field perpendicular to the plane of the figure (not shown).

far away from the oscillating charges we described in the last paragraph, they become equivalent to an oscillating electric dipole and the calculation of their effects is simplified. The electric field sent out by such an oscillator is shown for successive times in Fig. 263(a), (b), (c), (d). This electric field will be accompanied by a magnetic field perpendicular to the plane of the paper and varying in a similar way.

Standing Electric Waves

33. In Par. 18, we discussed standing waves on a chain as an example of interference between two trains of waves. Let us consider whether some such effect might not be found with elec-

trical waves. If one end of a wire is suddenly charged negatively, *i.e.*, given an excess of electrons, an electric field is set up directed along the wire toward that end. The excess electrons constituting the charge flow out through the wire until they are uniformly distributed and the electric field directed along the wire has disappeared. (It must be remembered that electrons move in a direction opposite to the electric field.) In considering a case like this in static electricity, we assumed that the uniform distribution took place instantaneously or that we were interested in the situation only after equilibrium had been established. Now we

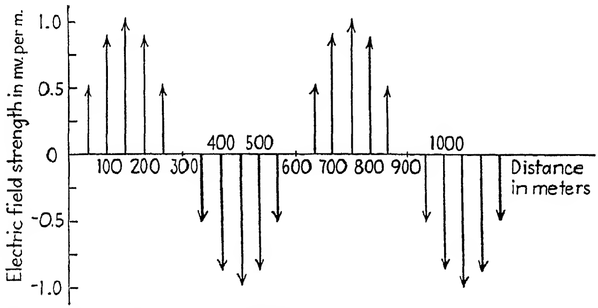


FIG. 264.—The electric field strength at some particular instant at successive points along the direction of propagation of a radio wave. The wave is assumed plane polarized, of 600 m. wave length and 1.0 millivolt per meter amplitude.

wish to consider what happens if the sign of the charge is changed before its field has spread down the wire and taken the charge with it. Such a change reverses the direction of the field and the motion of charges resulting from the field. Thus the field at the far end of the wire may be directed outward and the electrons moving back toward the near end at the time when the charge on the near end has been reversed so that the field close to it is directed inward and the electrons are moving outward. If the charge is repeatedly reversed, then a wave of alternating electric field strength is propagated down the wire. The corresponding electron motion is a wave of alternate compressions and rarefactions of the electrons. This wave travels down the wire, and since the electrons cannot escape at the end, the wave will be reflected. If the length of the wire is an integral number of half wave lengths, standing waves will result. When this condition has been set up, there will be certain points along the wire where

there is no compression or rarefaction, where the density of the electrons remains constant even though the electrons themselves may be moving back and forth. Such points may be called nodes of electron density since the electron density does not change with time. In the region near the end of the wire, on the other hand, electrons will alternately pile up and pull away so that the density of the electrons in this region varies between wide limits even though there can be no motion of the electrons through the end of the wire. In terms of electron density, then, the end of the wire will be an antinode and there will be other similar points along the wire. At such an antinode of electron density, there is alternately an excess of electrons (a resultant negative charge) and deficiency of electrons (a resultant positive charge). If two identical wires are connected to two identical condenser plates whose charges are oscillating with the same period but are always opposite in sign (*i.e.*, 180° out of phase), then the charges at the antinodes of the two wires will always be opposite and a small current will flow through a neon lamp connected between them. But at the nodes, there is never any excess charge and the lamp will not light. By determining the position of the nodes in such a system, the wave length of the electric wave can be measured. If the period of the oscillation of the charges is known, the speed of propagation of the waves can be determined since $V = \lambda/T$ (see Par. 10).

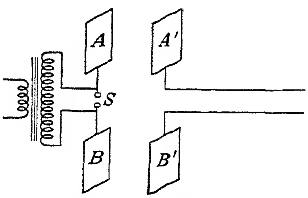


FIG. 265.—A Hertzian oscillator producing standing electric waves on two wires.

34. We know that V, the velocity of propagation of these waves, must be very large. Therefore to get standing waves short enough to be measured in the laboratory, T must be very small. No sort of mechanical charger will reverse the charges on condenser plates fast enough. What we do is to use a circuit like that shown in Fig. 265. The plates A and B form a condenser of very low capacity. They are charged by an induction coil or transformer until the voltage is high enough for a spark to jump across the gap S. When this spark jumps, it completes the oscillatory circuit ASB of very low capacity and inductance and therefore of very short period. The oscillating charges on A and B induce opposite charges on A' and B'. These of course

also oscillate, with the same period, and send out waves along the wires which are reflected at the ends of the wires and produce standing waves.

35. A similar experiment more difficult to perform shows the standing waves of electric and magnetic field strength produced in space by a short period electric oscillator sending out waves that are reflected back on themselves by a metallic surface.

36. Almost exactly a hundred years ago, Joseph Henry sent an oscillatory current from a discharging condenser through a wire on one side of Nassau Hall in Princeton and detected the induced effect in another wire 250 ft. away on the other side of the building. He even went so far as to suggest that this effect was transmitted by some sort of undulatory disturbance. Thus he anticipated modern radio in both practice and theory. Yet we can hardly call Henry the father of radio. Too long a gap intervened between his work and the work of Maxwell and of Hertz who really laid the basis for present-day radio.

The Velocity of Electromagnetic Waves.

37. It is unfortunate that Maxwell's work is a little too mathematical for us to consider, for it constituted one of the great intellectual triumphs of the nineteenth century. We can only say that he generalized the laws of electromagnetism in mathematical form. These laws expressed as four differential equations are still the foundation from which any deep study of electric and magnetic fields must start. From them, Maxwell deduced the remarkable conclusion that an electromagnetic disturbance is propagated through space with a velocity that is just equal to one over the square root of the product of the dielectric constant and the permeability for empty space, $1/\sqrt{\epsilon_0\mu_0}$. He also showed that such a disturbance would satisfy the mathematical equations characteristic of waves. Maxwell's great treatise on electricity and magnetism was published in 1873, but it was not until the work of Hertz in the eighties that his predictions were directly verified. Hertz was able to produce electromagnetic waves, to measure their velocity and wave length, to show that they behaved as transverse waves, and generally to establish the correctness of Maxwell's predictions.

38. We have said that Maxwell showed that electromagnetic waves would be propagated with a velocity equal to $1/\sqrt{\epsilon_0\mu_0}$.

From laboratory measurements of ϵ_0 and μ_0, this predicted velocity is found to be approximately 3×10^8 m./sec. (= 186,000 mi./sec). This is an enormously great speed, as we anticipated, but not only is it big, it is exactly equal to the only other known speed of comparable magnitude, a speed that had been known for many years before electromagnetic waves were ever heard of, the speed of light. That light was a wave of some sort had been known for a good many years before Maxwell's time. That it was an electromagnetic wave was a new and revolutionary idea.

39. From our modern point of view about the structure of matter, it is not surprising that light is an electromagnetic wave. Now that the speed of propagation of such a wave has been given, let us review the nature of the waves sent out from various oscillating electrical systems. In a slowly oscillating system, for example, one with a capacity of 250 microfarads and an inductance of about 100 henries, the period is of the order 1 sec., and therefore the wave length is about 186,000 mi. ($\lambda = VT$), and we could hardly expect to notice any wave characteristics. In ordinary 60-cycle alternating current, the wave length is 3,100 m. If we used a circuit consisting of an air-coil inductance and a small parallel-plate air condenser, we would have an inductance of perhaps 10 millihenries and a capacity of perhaps 1/10,000 microfarad. This would give a period of $2\pi \times 10^{-6}$ sec. or a wave length of approximately 2,000 m. This is in the ordinary wireless range. Now consider the smallest electrical system that we can think of, an atom. It consists of electrons and a positively charged nucleus. If the thermal motion of atoms makes them collide with each other, this electrical system must be disturbed and might be expected to send out electromagnetic waves. It is possible to take the dimensions of an atom and make a rough estimate of its equivalent inductance and capacity and therefore of the wave length that it might radiate. But this would be silly since we would be applying laws that were developed for large circuits to tiny electrical systems whose very existence can be determined only indirectly. Let us content ourselves by saying that disturbed atoms or "excited atoms," to use the technical expression, might be expected to give off electromagnetic radiation of very short wave length. Actually, the wave length of the part of this radiation that we perceive as light varies between 3,900 and $8{,}000 \times 10^{-10}$ m. The extremely short wave length

of this radiation gives it properties that are different in many ways from those of longer electromagnetic waves. In particular, the fact that we perceive it directly with our eyes has made it very much more familiar than other forms of radiation. Naturally, it was studied long before electromagnetic waves were known, and it is still desirable to treat it as a more or less separate

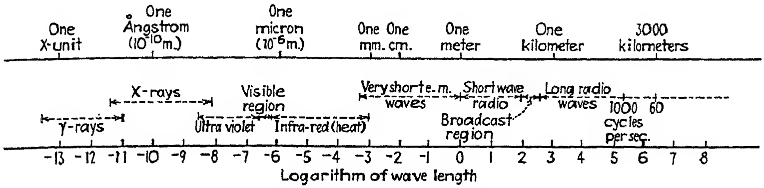


FIG. 266.—The complete electromagnetic spectrum plotted on a logarithmic scale. See Chap. XXVI for further discussion of this subject.

subject. A brief treatment of some of the principal properties of light will be given in the next chapter.

SUMMARY

In nature, the most important means of transmitting energy is by trains of waves. Energy from the sun is communicated to the earth by electromagnetic waves. A train of waves is a periodic disturbance propagated with a finite velocity. The most familiar waves are those on the surface of water or the elastic waves in a spring or string. In a transverse wave, the direction of the disturbance is perpendicular to the direction of propagation; in a longitudinal wave, the direction of the disturbance is parallel to the direction of propagation. The simultaneous variation in time and space characteristic of a train of waves is described by the wave equation

$$s = s_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right)$$

where s is the value of the disturbance at the time t at a distance x along the direction of propagation. The disturbance s may be an actual mechanical displacement, a pressure, an electric field strength, or some other varying physical quantity. In the preceding equation, s_0 is called the amplitude, T the period, and λ the wave length. The frequency f or ν of a wave is $1/T$. The

amount of energy transmitted by a train of waves is proportional to the square of the amplitude.

Sound waves are compressional longitudinal waves. The pitch of a pure tone depends on the frequency. If either the source or observer of a train of waves is in motion, the effective frequency is altered. This is known as the Doppler effect.

The resultant effect of two trains of waves at the same point depends on the phase relations. Such waves are said to interfere. The interference between the waves going one way down a string and the reflected waves coming back produces standing waves in which certain points called nodes have no resultant motion.

A transverse wave is polarized if the disturbance is not merely confined to the plane perpendicular to the direction of propagation but has a definite direction in that plane.

A finite time is required for the establishment of an electric or magnetic field in space so that an oscillating electric charge sends out waves of electric and magnetic field strength varying periodically in magnitude and direction. Standing electric waves can be produced both in wires and in space. Maxwell predicted that electromagnetic waves should be propagated with a velocity equal to $1/\sqrt{\epsilon_0\mu_0}$, a velocity that proves to be equal to the velocity of light. Light is an electromagnetic wave like a radio wave but of very much smaller wave length.

ILLUSTRATIVE PROBLEMS

1. A wave with a wave length of 0.75 m. and a period of 0.1 sec. travels down a string. If the displacement at $x = 1$ m. is 1 cm. when $t = 0$, what will be the displacement of a point 2 m. down the string after 1.01 sec?

Equation (3), page 533, gives the displacement of a point on a string down which a wave travels. Thus,

$$s = s_0 \sin 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right)$$

where s = the displacement at a distance x from the origin at time t.

s_0 = the amplitude of the wave.

T = 0.1 sec. = the period.

λ = 0.75 m. = the wave length.

Initially, $t = 0$, $x = 1$ m., $s = 1$ cm.

$$1 \text{ cm.} = s_0 \sin\left(-2\pi \text{ rad. } \frac{1 \text{ m.}}{0.75 \text{ m.}}\right) = s_0 \sin(-120°) = -s_0(0.866)$$

$$s_0 = \frac{-1 \text{ cm.}}{0.866} = -1.15 \text{ cm.} \qquad \text{so that}$$

$$s = (-1.15 \text{ cm.}) \sin 2\pi \text{ rad.} \left(\frac{t}{0.1 \text{ sec.}} - \frac{x}{0.75 \text{ m.}}\right)$$

After 1.01 sec., 2 m. down the string

$$\begin{aligned} s &= -1.15 \text{ cm. } \sin 2\pi \text{ rad.} \left(\frac{1.01 \text{ sec.}}{0.1 \text{ sec.}} - \frac{2 \text{ m.}}{0.75 \text{ m.}}\right) \\ &= -1.15 \text{ cm. } \sin 2\pi \text{ rad. } (10.1 - 2.7) = -1.15 \text{ cm. } \sin 2\pi \text{ rad. } 7.4 \\ &= -1.15 \text{ cm. } \sin 2\pi \text{ rad. } 0.4 = -1.15 \text{ cm. } \sin 2.51 \text{ rad.} \\ &= -1.15 \text{ cm. } \sin 144° = -1.15 \text{ cm. } \sin 36° = -1.15 \text{ cm. } 0.588 \\ &= -0.676 \text{ cm.} \end{aligned}$$

2. Two cars pass each other on the highway traveling in opposite directions. The horn of car A emits 608 vibrations per second which the driver of car B hears as 673 vibrations per second. The horn of car B emits 529 vibrations per second which the driver of car A hears as 581 vibrations per second. What is the speed of each car?

Equation (10), Par. 16, gives as the frequency heard when the ear and the source both approach each other

$$n'' = n\frac{V + v_e}{V - v_s} \tag{21}$$

where n'' is the frequency heard and n the frequency emitted. V is the velocity of sound, v_e the velocity of the ear, and v_s that of the source.

First let car A be the source

n_A = 608 vibrations/sec. = the frequency emitted by car A.
n''_B = 673 vibrations/sec. = the frequency heard by the driver of car B.
V = 1,129 ft./sec. = the velocity of sound at 20°C.
$v_A = v_s$ = the velocity of car A as source.
$v_B = v_e$ = the velocity of car B as ear.

Substituting these values in Eq. (1) gives

$$673 \text{ vibrations/sec.} = 608 \text{ vibrations/sec.} \frac{1{,}129 \text{ ft./sec.} + v_B}{1{,}129 \text{ ft./sec.} - v_A}$$

from which

$$\begin{aligned} (1{,}129 \text{ ft./sec.} - v_A)1.107 &= 1{,}129 \text{ ft./sec.} + v_B \\ 1{,}250 \text{ ft./sec.} - 1.107v_A &= 1{,}129 \text{ ft./sec.} + v_B \\ v_B &= 121 - 1.107v_A \end{aligned} \tag{22}$$

Secondly, let the car B be the source.

n_B = 529 vibrations/sec. = the frequency emitted by car B.
n''_A = 581 vibrations/sec. = the frequency heard by the driver of car A.
$v_B = v_s$ = the velocity of car B as source.
$v_A = v_e$ = the velocity of car A as ear.

Substituting these values in Eq. (21) gives

$$581 \text{ vibrations/sec.} = 529 \text{ vibrations/sec.} \frac{1{,}129 \text{ ft./sec.} + v_A}{1{,}129 \text{ ft./sec.} - v_B}$$

from which

$$(1{,}129 \text{ ft./sec.} - v_B)1.098 v_B = v_A + 1\ 129 \text{ ft./sec.}$$
$$111 \text{ ft./sec.} - 1.098 v_B = v_A$$

Substituting the value of v_B from Eq. (2) gives

$$v_A = 111 \text{ ft./sec.} - 1.098(121 \text{ ft./sec.} - 1.107 v_A)$$
$$v_A(1 - 1.215) = 111 \text{ ft./sec.} - 133 \text{ ft./sec.} = -22 \text{ ft./sec.}$$
$$v_A = \frac{-22 \text{ ft./sec.}}{-0.215} = 102 \text{ ft./sec.}$$

Substituting this value of v_A in Eq. (22) gives

$$v_B = 121 \text{ ft./sec.} - 1.107 \times 102 \text{ ft./sec.} = 8 \text{ ft./sec.}$$

PROBLEMS

1. A wave with an amplitude of 5 cm. and a frequency of 10 vibrations per second travels down a string with a velocity of 15 m./sec. What will be the displacement 10 m. down the string after 1 sec.?

2. A wave making 15 vibrations per second travels down a string with a velocity of 20 m./sec. What is the difference in phase of two points 1 m. apart?

3. A wave with a wave length of 1.4 m. travels down a string with a velocity of 5 m./sec. If the displacement 1 m. down the string is 5 cm. when $t = 0$, what will be the displacement 9 m. down the string after 1 sec.?

4. A wave having an amplitude of 2 cm., a wave length of 1 m., and a frequency of 30 vibrations per second travels down a stretched rope. Initially, a point on the rope has a displacement of -1 cm. After $\frac{5}{12}$ of a period, what is the displacement of this point, and how far down the rope is the nearest point having the same displacement?

5. A continuous wave with an amplitude of 4 cm. and a frequency of 17 vibrations per second is sent down a chain with a velocity of 11 m./sec. The remote end of the chain is fixed. The displacement at a particular instant when the wave first passes a point 3 m. from the fixed end of the chain is zero. Find the displacement at this point after $\frac{5}{3}$ sec.

6. The world's record of 10.3 sec. for 100 m. was set by Eddie Tolan at Los Angeles, Aug. 1, 1932. How much error is introduced in this record by having the starting gun fired at a distance 20 ft. from the runners?

7. An automobile with the horn blowing at 500 vibrations per second approaches a cliff. If the driver hears the pitch of the horn increased by $\frac{6}{5}$ on reflection from the cliff, what is the speed of the car?

8. A train roars through the station at 90 mi./hr. blowing its whistle. If a man on the platform hears the pitch of the whistle drop 142 vibrations per second as the train passes, what is the frequency of the whistle?

9. Two cars pass on the highway traveling in opposite directions. The horn of one car is heard by the driver of the other. Show that the relative change of pitch when the cars pass is given by

$$\frac{\Delta n}{n} = \frac{2V(v_e + v_s)}{V^2 - v_s^2}.$$

10. A car going 70 mi./hr. in overtaking another car blows its horn at 600 vibrations per second. If the car overtaken is going 50 mi./hr., what drop in pitch will the driver hear as the other car passes?

11. How many nodes are there on a string 3 m. long if one end moves up and down with simple periodic motion twenty times a second and the other end is fixed? The waves travel 48 m. a second on the string.

12. Show from the units and numerical values that we have used for ϵ_0 and μ_0 that $1/\sqrt{\epsilon_0\mu_0}$ is a velocity of 3×10^8 m./sec.

CHAPTER XXVI

LIGHT

1. In the last chapter we discussed the existence of electromagnetic waves and saw that they were propagated with the speed of light. We saw that this implied that light itself might be an electromagnetic wave. In the present chapter we want to give a brief discussion of the properties of light in general and in particular to show that these properties are explicable on the assumption that light consists of electromagnetic waves. Since the first property of light which we have mentioned is its velocity, our first concern will be to see how this is measured.

The Velocity of Light

2. The first recorded attempt to measure the velocity of light was made by Galileo, who tried to measure the time required for the light of a lantern to travel from one hilltop to another. His experiments gave inconsistent results, and he rightly concluded that the velocity of light was too great to be measured by the crude method he used.

3. Evidently, if light travels very rapidly, a measurement of its speed must use either very great distances or very precise methods of measuring time. So it is not surprising that no significant results were obtained until an astronomical method was used that measured the time required for light to cross the earth's orbit. The observations were made and the velocity of light calculated from them by a Danish astronomer Roemer working in Paris in the last half of the seventeenth century. They are clearly described by Huygens in his "Traité de la Lumière" published in 1690, from which we shall quote in free translation. After discussing the possibility of measuring the velocity of light by observations of the eclipses of the moon, and showing that they are accurate enough to prove that the velocity must be many times greater than that of sound, but not accurate enough to give an actual value for the velocity of light, Huygens continues as follows:

4. It may seem astonishing to assume a velocity which may be a hundred thousand times greater than that of sound. According to my observations sound travels about 180 fathoms (about 1,100 ft.) in a time of one sec. or one pulse beat. But such an assumption does not appear to be impossible; for it is concerned not with the advance of a body itself with this great a velocity but with the transfer of a state of motion from one body to another. Consequently in thinking about these things I have no difficulty in supposing that the propagation of light requires time for, on this view, all the phenomena can be explained while on the opposite view (infinite velocity) they are entirely incomprehensible. . . . This supposition of mine with regard to the velocity of light now appears to have been verified by an ingenious demonstration by Roemer which I will describe. . . . His conclusions, like mine, are based on astronomical observations and not only prove that light requires time for its transmission but show how much time is needed and that its velocity is six times as large as I have suggested.

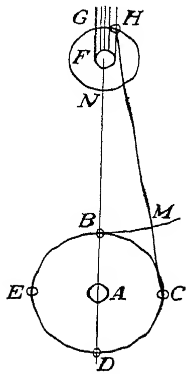


FIG. 267.—Roemer's method for determining the velocity of light by observations on the eclipses of one of Jupiter's satellites.

5. Roemer uses the small moons which revolve around Jupiter and which often pass into its shadow. His argument is as follows. Let A (Fig. 267) be the sun, $BCDE$ the annual orbit of the earth, F Jupiter, and GN the orbit of the innermost satellite, since this one is better suited to this investigation than any of the other three because of its short period of revolution. Let G be the point where the satellite enters Jupiter's shadow and H be the point where it emerges.

6. Suppose that when the earth is at B . . . the satellite is observed emerging from the shadow; if the earth remained motionless, the next emergence would be seen in 42.5 hr., for this is the time the satellite needs to make a circuit of its orbit and return to its previous position. . . . And if the earth stayed at B for, say, 30 revolutions of the satellite, then it would appear out of the shadow once more after exactly thirty times 42.5 hr. But in fact the earth in such a time would move to C, farther away from Jupiter, so that, if light needs time for its transmission, the appearance of the satellite would be observed later at C than it would have been at B. It would be necessary to add to 30×42.5 the time needed for light to travel the distance MC, the difference between the distances CH and BH. Similarly when the earth is on the other side of its orbit moving from D to E, approaching Jupiter, the

entrances of the satellite into the shadow G should be observed earlier at E than if the earth had stayed at D.

7. From a great number of observations of these eclipses made during ten consecutive years these differences are found to be considerable, as much as 10 min. or more, and it is possible to conclude that light takes in the neighborhood of 22 min. to traverse a distance equal to the diameter of the earth's orbit, twice the distance from here to the sun.

8. The movement of Jupiter in its orbit while the earth is moving from B to C or from D to E is included in the calculations, and it is also necessary to see that the observed effects could not be explained by irregularities in the motion of the satellite or by its eccentricity.

9. Huygens then proceeds to a calculation of the velocity of light, arriving at a value equivalent to about 212,000 km./sec. More recent observations on Jupiter's satellites have shown that the time of transit of light across the earth's orbit is about 16 min. instead of 22. The diameter of the earth's orbit has also been

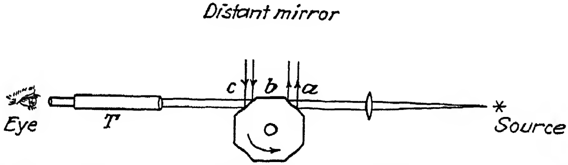


FIG. 268.—Michelson's method for measuring the velocity of light.

found somewhat greater than that assumed by Huygens, which accounts for the rest of the discrepancy between this value of Roemer's and the modern value of about 299,790 km./sec.

10. Between the time that Huygens wrote this discussion of Roemer's observations and the present, there have been perhaps 20 different measurements of the velocity of light which were of sufficient accuracy to be valuable. The only ones that we shall discuss are the recent measurements of Michelson, using a method originated by Foucault. In this method, a beam of light is reflected from a rotating mirror to a distant fixed mirror from which it returns again to the rotating mirror. In the meantime, this mirror has turned so that the light is reflected to a point different from its original source. A schematic diagram of Michelson's apparatus is given in Fig. 268. In his experiment, light from the sharply defined source struck the face a of the eight-sided mirror. From there, it traveled about 23 mi. to a fixed

plane mirror on a distant mountain top. Returning by a slightly different path, it struck the face c of the eight-sided mirror and was reflected into the telescope T, where a sharply focused image of the source could be observed. If the mirror were rotating, pulses of light would be sent to the distant mirror every time a face of the rotating one was in the position of a. The face c would move somewhat in the time that the light pulse took to travel to the plane mirror and back. In general, the new position of the mirror would not send the reflected beam into the telescope, but for the right speed of rotation, the face b of the mirror would come around into approximately the position previously held by c so that the reflection would once more be into the telescope. The exact position of the image could be measured and compared with that when the mirror was stationary. By this measurement of the position of the image, by a measurement of the speed of rotation of the mirror, and by a measurement of the distance between the rotating mirror and the plane mirror on the mountain top, the velocity of light could be determined. The distance between the rotating mirror on Mt. Wilson and the plane mirror on Mt. San Antonio was determined by the Coast and Geodetic Survey as 35,426.3 m. and was considered accurate to within 0.1 m. The speed of rotation of the mirror was determined within 1 part in 100,000, which was equivalent to measuring the time for the light to travel to Mt. San Antonio and back (about $\frac{1}{5000}$ sec.) to an accuracy of two thousand-millionths of a second. The final result of all these measurements when corrected for the effect of the air was to give $299{,}796 \pm 4$ km./sec. for the velocity of light in vacuum.

11. A few years later, Michelson and Pease (1930–1931) obtained a value of 299,774 km./sec. for the velocity of light in a mile-long pipe from which most of the air had been pumped. This was also supposed to be a very accurate determination, and the discrepancy between the two values has never been satisfactorily explained.

Straight-line Propagation, Reflection, and Refraction

12. These are the most familiar properties of light. We shall go over them briefly. That light travels in straight lines is obvious from the geometrical sharpness of the shadows cast by any light source that is even approximately a point, *e.g.*, the

sun, an arc light, or a single incandescent bulb more than a few feet away. At first sight, straight-line propagation appears incompatible with the wave theory, but we shall see later that it is not exactly true and that its approximate truth is the result of the very small wave length of light. The property of straight-line propagation enables us to represent the behavior of beams of light by drawing straight lines which we call rays.

13. Thus we can discuss the law of reflection in terms of the behavior of a beam of parallel rays of light, *i.e.*, a beam whose cross section is constant. The direction of such a beam is specified by the direction of any ray in it. For example, the incident beam of light in Fig. 269 makes an angle $\left(\frac{\pi}{2} - i\right)$ with the surface S and an angle i with a line perpendicular to that surface. This latter angle is called the angle of incidence. The angle i' that

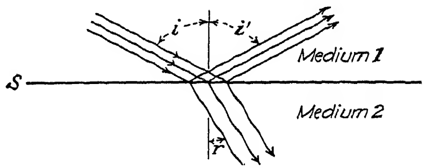


Fig. 269.—The reflection and refraction of rays of light at the surface between two mediums. i is the angle of incidence, i' the angle of reflection and r the angle of refraction.

the reflected beam makes with this same line is called the angle of reflection. Simple measurements show that:

The angle of reflection is always equal to the angle of incidence and the reflected beam is in the plane determined by the incident beam and a line perpendicular to the reflecting surface.

14. If light falls on a surface separating two transparent mediums such as air and water, some of it is reflected according to the foregoing law but some of it goes through the surface. The direction of the rays changes suddenly at the surface. For this reason, the light is said to be refracted. The rays in the second medium are called refracted rays, and the angle they make with the perpendicular to the surface is called the angle of refraction. The law of refraction is not so simple as the law of reflection, but nevertheless has been known for a very long time. It is stated as follows:

When a beam of light passes through the surface of separation of two transparent mediums, the refracted beam is in the plane determined by the incident beam and the perpendicular to the surface of separation. The ratio of the sine of the angle of incidence to the sine of the angle of refraction is a constant, known as the index of refraction.

In mathematical language,

$$\frac{\sin i}{\sin r} = n \tag{1}$$

where n is the index of refraction and depends not only on the nature of the two mediums but on the wave length. The values of the index of refraction given in tables are usually for the passage of light from a vacuum into the substance and for a particular wave length, frequently the yellow light of sodium vapor (wave length = $5{,}893 \times 10^{-10}$ m.). The index of refraction for the passage of light from one medium to another is the index of the second medium relative to vacuum divided by the index of the first medium relative to vacuum.

TABLE 19.—INDICES OF REFRACTION

Air	1.000294
Benzene	1.495
Carbon tetrachloride	1.46
Glass, silicate flint	1.5794
Borosilicate flint	1.5503
Heavy flint	1.9625
Silicate crown	1.4782
Borosilicate crown	1.4944
Heavy barium crown	1.6130
Lubricating oil (automobile)	1.50
Water	1.33

Total Reflection

15. The index of refraction varies roughly with the density so that light going from a less dense to a more dense medium is refracted toward the perpendicular, and vice versa. If light is passing from a dense to a rare medium, from water to air, for instance, an interesting situation arises as the angle of incidence increases. Since the refracted beam is bending farther away from the perpendicular than the incident beam (Fig. 270), it soon makes an angle of 90° with the perpendicular, *i.e.*, it grazes the

surface of separation. For any further increase in the angle of incidence, the refracted beam cannot get through the surface and all the light is entirely reflected as shown in Fig. 271. This phenomenon is known as total reflection. In terms of Eq. (1), total reflection occurs when the $\frac{\sin i}{n} = \sin r$ becomes greater than one. Mathematically it is impossible to have an angle whose sine is greater than one, and therefore no angle of refraction can exist that satisfies this requirement. This confirms the physical observation that there is no refracted beam under these conditions.

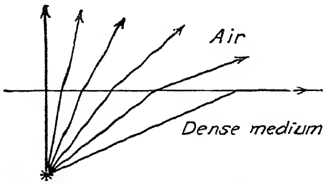


FIG. 270.—The refraction of light rays emerging from a dense medium into air showing the beginning of total reflection.

Geometrical Optics

16. The whole subject of lenses, cameras, telescopes, microscopes, etc., can be built up on the basis of the three simple laws: straight-line propagation, the law of reflection, and the law of refraction. Readers of this book have undoubtedly had some elementary study of this part of the subject. A repetition of this elementary treatment does not seem worth while, and we have not time for a more advanced consideration of these topics. The student is advised to refresh his mind as to the meaning of such

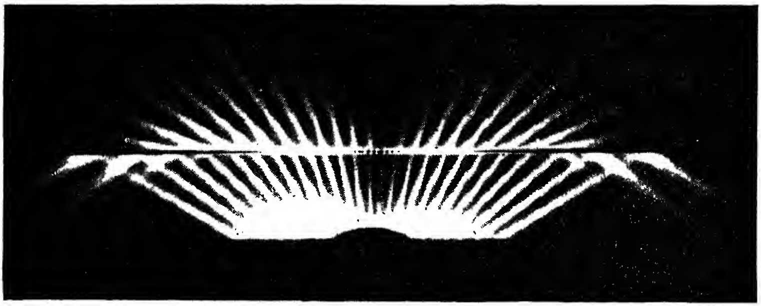


FIG. 271.—Photograph showing total reflection. The light source is in water at the bottom of the figure. The last two rays on each side are totally reflected at the surface of the water.

terms as focus, real image, virtual image, and convergent and divergent lenses by reviewing his school text or by looking up these terms in a good dictionary.

The Wave Point of View

17. If light really does consist of waves, then the preceding laws must be explicable in terms of waves. We have already suggested that the property of apparently straight-line propagation is the result of the small wave length. Suppose we argue by analogy. We have stated and shall presently prove that the wave length of light is about 5×10^{-7} m. The obstructions whose sharp shadows we observe are at least half a millimeter wide, about a thousand wave lengths. If we put a breakwater 20,000 ft. long in the path of sea waves 20 ft. long, we get a pretty definite shadow even though the waves are certainly not coming from a point source. To be sure, the edge of the shadow is not very sharp, but neither is that of the light shadow if it is examined sufficiently closely.

18. In order to consider the reflection and refraction of waves, we shall introduce an idea originated by the Dutch physicist Huygens at the end of the seventeenth century. When a single wave travels down a string, it is clear that the immediate cause of the sideways displacement of a particular part of the string is not the original disturbance at the end of the string, which may have subsided long since, but the displacement of the particles of the string immediately adjacent to the part under consideration. Similarly, when a group of waves spreads out from the splash of a stone in a pond, new waves are continually forming in larger and larger circles long after the water has quieted down where the stone originally hit. The train of waves keeps itself going, leaving quiet water behind it, but continually stirring up the water in front of it. Evidently the front wave itself is a line of disturbance just as if a lot of tiny stones had been dropped simultaneously along the circle it forms. From this point of view, every point along the surface of a wave may be considered a new source of waves and the condition a moment later is the resultant of all the wavelets spreading out from the front of the old wave. Such a way of looking at waves is called Huygens' principle. It proves to be a fruitful method of attacking many problems of optics.

19. This point of view is illustrated by Fig. 272 where the semicircles represent wavelets traveling out from the circular wave front AB and forming the new wave front $A'B'$. This new wave front is obtained by drawing the surface that is tangent to the

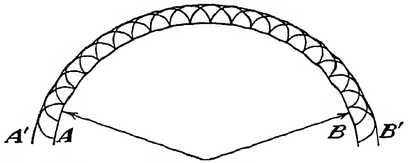


FIG. 272.—Huygens' principle. The wave front $A'B'$ is formed from all the wavelets originating on the wave front AB.

surfaces of all the wavelets. Mathematically, the new wave front is known as the envelope of the wavelets. That this is not merely an arbitrary mathematical construction can be shown by studying the waves in a water tank. If a series of parallel waves on a water surface strikes a straight barrier with a series of small openings in it, these openings will become new sources of waves

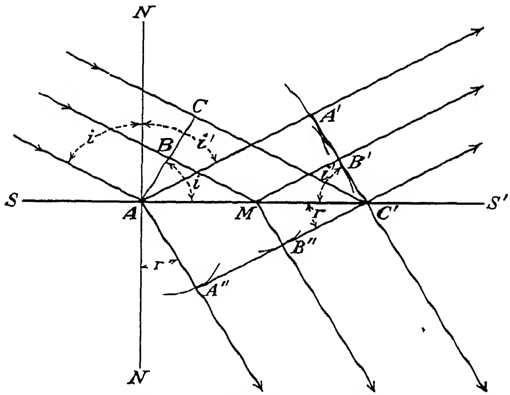


FIG. 273.—Derivation of the laws of reflection and refraction by a construction based on Huygens' principle.

and will each send out a system of circular wavelets. These wavelets are seen to merge a few wave lengths beyond the apertures and to form single waves parallel and similar to the original waves.

20. Now let us apply Huygens' principle to the problem of reflection of a plane wave from a plane surface. By a plane wave,

we mean a wave in which the points of equal phase at any instant lie in a plane. Such a wave corresponds to a beam of parallel rays, and the direction of propagation of the wave is perpendicular to the plane of equal phase. In Fig. 273, such a wave is shown striking the surface SS' at an angle of incidence i. The problem before us is to determine the angle of reflection by applying Huygens' principle. Draw ABC, the wave front the end of which, A, is just touching the surface. This wave front will reach the points on the surface from A to C' at successively later times. As the wave front reaches each of these points, it sets up a disturbance at the surface, and each of these points may be considered a new center from which a wavelet spreads out according to Huygens' principle. The part of this wavelet that travels back into the medium from which the incident wave comes contributes to the reflected wave. The part that travels down into the new medium contributes to the refracted wave. For simplicity, we shall consider the wavelets from only the three points, A where the plane wave first hits the surface, M where the mid-point B of the wave hits the surface, and C' where the other end of the wave hits the surface. If the wave is traveling with a velocity V, the wavelet from M will start out BM/V sec. after that from A and the wavelet from C' will start out CC'/V sec. after that from A. By the time that the wavelet is starting from C', that from A will have spread out into a circle of radius

$$AA' = V \times \frac{CC'}{V} = CC'$$

and that from M will have spread out into a circle of radius

$$MB' = V \times \left(\frac{CC'}{V} - \frac{BM}{V}\right);$$

but by similar triangles, $BM = \dfrac{CC'}{2}$ and therefore $MB' = \dfrac{CC'}{2}$. The new reflected wave front will be given by a line tangent to these circles and through the point C', *i.e.*, by the line $A'B'C'$. The angle of reflection i' is then the angle $A'AN = A'C'A$ and $\sin i' = AA'/AC'$. But the angle $CAC' = i$ and $\sin i = CC'/AC'$. Since $AA' = CC'$, $\sin i = \sin i'$ or $i = i'$ and we have proved the law of reflection by the use of Huygens' principle.

21. The law of refraction is derived by a similar construction based on the fact that the velocity of the waves in the second medium will be different. The surface of the refracted wave $A''B''C'$ in Fig. 273 is constructed on the assumption that the velocity V' in the second medium is less than the velocity V in the first. The radius of the wavelet AA'' will then be

$$AA'' = V' \times {}^{CC'}$$

and that of the wavelet from M will be $MB'' = V' \times CC'/2V$. The angle of refraction is $A''AN'$ and this is equal to the angle $A''C'A$. But the sine of $A''C'A$ is AA''/AC'. Therefore the ratio of the sine of the angle of incidence to the sine of the angle of refraction is

$$\frac{\sin i}{\sin r} = \frac{CC'/AC'}{AA''/AC'} = \frac{CC'}{AA''} \quad \frac{CC'}{V'CC'/V} \quad \frac{V}{V'}.$$

i.e.

$$\frac{\sin i}{\sin r} = n = \frac{V}{V'} \tag{2}$$

22. Thus we have proved not only that the law of refraction is compatible with the wave theory but that the index of refraction is not merely an arbitrary constant, but is the ratio of the velocities of propagation of light in the two mediums. The velocity of light has been measured directly in a few mediums besides air and the foregoing relation confirmed.

Interference of Light

23. We have seen that the most familiar properties of light are not inconsistent with the theory that light is an electromagnetic wave, but it is evident that the properties of light we have presented so far would hardly have suggested the wave theory. The next phenomena that we shall discuss are of such a nature that they have never been satisfactorily explained except in terms of a wave theory. The student should perform the experiments himself or see them demonstrated. Here they will be described in terms of the wave theory.

24. *Young's Experiment.* Suppose that light from the sun or any bright small source falls on two tiny pinholes or slits in a

card or screen. Two narrow trains of waves will emerge on the other side of the slits. Let these trains of waves or beams of light fall on a white screen. Suppose that the distance apart of the slits as shown in Fig. 274 is d and the perpendicular distance from M, the middle of this line, to the point O on the screen is L. Then if the two trains of waves leave the slits S and S' in phase they will reach O in phase since it is equidistant from S and S'. But for any other point on the screen, the distance to S is different from the distance to S'. If this difference is half a wave length, three half wave lengths, or any odd number of half wave lengths, the trough of the wave from S will coincide with the crest of the wave from S' and the resultant effect will be zero.

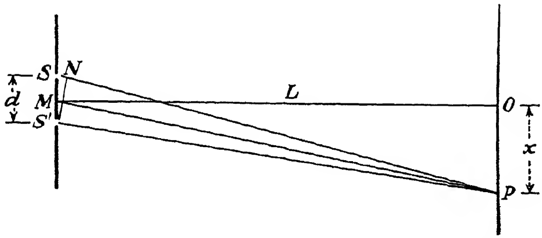


FIG. 274.—Young's interference experiment. A beam of light from the left falls on the slits S and S'. Light from S and S' combines to give alternate bright and dark interference fringes on the screen above and below O.

If, on the other hand, the difference in the distances is a whole wave length or an integral number of whole wave lengths, the crests of the two waves (and the troughs) will arrive together and there will be a maximum effect. If d and L are large compared with the wave length the path difference between the two beams can be easily computed. Take any point P a distance x below O. Then draw the lines PS, PM, and PS'. The difference between PS and PS' is the difference in path length of the light getting to P from the two holes. With P as center and PS' as radius draw an arc cutting PS at N. If SS' is small compared to OM, this arc is practically a straight line, the angle $NS'S$ is equal to the angle OMP and the triangles $NS'S$ and OMP are similar. Therefore NS/SS' equals OP/PM, but PM is very nearly equal to $OM = L$. Therefore, to a first approximation

$$\frac{NS}{SS'} = \frac{OP}{OM} \quad \frac{x}{L}$$

or

$$\frac{PS - PS'}{d} = \frac{x}{L}, \qquad \text{or} \qquad x = \frac{L}{d}(PS - PS').$$

The value which x must have in order to have a path difference of $\lambda/2$ between the waves coming from the two slits is obtained by setting $PS - PS' = \lambda/2$. Accordingly, the distance of the center of the first dark interference band below O is given by

$$x_1 = \frac{L}{d}\frac{\lambda}{2} \tag{3}$$

and of the next by

$$x_2 = \frac{L}{d}\frac{3\lambda}{2}.$$

Evidently there will be a dark band whenever

$$x = \frac{L}{d}\left(\frac{2n - 1}{2}\right)\lambda \tag{4}$$

where n is a whole number. The distance between any two adjacent dark bands is

$$x_2 - x_1 = \frac{L}{d}\left(\frac{3\lambda}{2} - \frac{\lambda}{2}\right) = \frac{L}{d}(\lambda). \tag{5}$$

If the position of these bands can be determined and the distances L and d measured, the wave length of the light can be calculated.

Newton's Rings

25. Probably the only example of the interference of light that is universally familiar is the phenomenon known as Newton's rings. These colored rings or bands occur whenever light is reflected at both the upper and lower surfaces of a thin layer of some transparent substance. The most familiar example is the reflection of light from a thin film of motor oil on a wet road. Because the oil does not mix readily with the water on the road's surface, the underside has a definite smooth surface as well as the upper side. The light is partially reflected and partially refracted at the upper surface. Much of that which is refracted is reflected at the lower surface and comes back through the upper surface

into the air again. Here it reunites with the light reflected from the upper surface. The difference of phase between the rays coming from the same point on the top surface, the one directly reflected and the other after passing twice through the film of oil, will depend on the thickness of the film and on the angle at which the ray has gone through it. If the light falling on the film is all the same color, there will be alternate bright and dark bands corresponding to path differences of even or odd numbers of half wave lengths. If the incident light is white, containing all the wave lengths of the visible spectrum, the dark bands for different

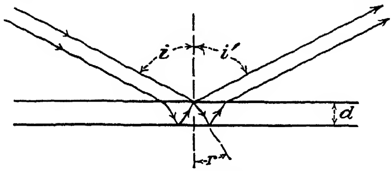


FIG. 275.—The formation of Newton's rings by interference between the light reflected from the upper and lower surface of a thin film.

colors come at different places so that the resulting light always lacks at least one wave length and is no longer white but colored.

Diffraction

26. Suppose that the slits in Young's experiment are gradually widened by moving their outer edges farther and farther away. The distance from one of the slits to any particular point on the screen can no longer be considered constant. It will differ for different parts of the slit. It will no longer be possible to decide whether there will be constructive or destructive interference at a given point on the screen by comparing the distances PS and PS'. We must also consider the phase relations of the wavelets coming from different parts of the same slit. This is somewhat too complicated mathematically for us to consider quantitatively. Even for the light coming through a single slit, we shall not attempt a mathematical treatment, but we shall describe the results of experiment which are confirmed by mathematical analysis. This type of phenomenon which results from the interference of different parts of the same light wave is called diffraction.

27. According to Huygens' principle, every point on the wave front may be considered a new center of disturbance. Thus when

a plane wave passes through a slit, each point in the slit may be considered a new source of light. If this is true, it is a surprising fact that light does travel in straight lines. It can be shown, however, that the various little wavelets interfere with each other so that there is no illumination anywhere in the geometrical shadow except near its edge. Near the edge of the geometrical shadow, interference between the wavelets results in narrow bands of destructive and constructive interference which are called diffraction bands. These bands are colored if white light is used since the positions of the diffraction minima are different for different wave lengths. Their width is inversely proportional to the width of the slit used. If the slit is made very narrow, it is impossible to distinguish any sharp edge to the shadow, and the illumination spreads over a wide area bearing no relation to the geometrical image of the slit.

28. If instead of a slit a small hole is used, diffraction rings are observed. If the slit or hole is replaced by a wire or disk, similar patterns are found. Perhaps the most striking diffraction effect that occurs is the bright spot that is found at the center of the geometrical shadow of a circular disk, but it is hard to get experimental conditions right to show this. All these interference and diffraction effects are explicable in terms of a wave theory but have never been explained successfully by any other theory.

The Diffraction Grating

29. We pointed out that Young's experiment could be used to measure the wave length of light. This is true of other interference and diffraction experiments. But the instrument that is used most to get measurements of these wave lengths is known as a diffraction grating. In its simplest form as shown in Fig. 276, it consists of a very large number of very narrow parallel slits close together. Light from some source, preferably a line source such as may be formed by an illuminated slit, is made into a parallel beam of plane waves by a convergent lens. This plane wave of light falling perpendicularly on the grating sends out new Huygens' wavelets from each slit. These wavelets all start in phase. Another convergent lens put on the other side of the grating will bring these waves to a focus. For the direction perpendicular to the grating, the wavelets from all the slits will

unite in phase for all wave lengths and will form an uncolored image of the light source. But there will also be other directions in which the wavelets from the different slits will combine to give

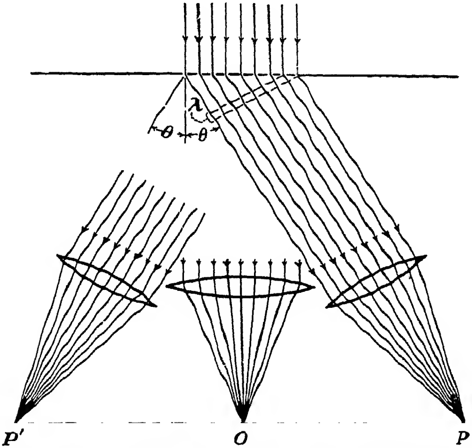


FIG. 276.—The formation of images by light after passing through a diffraction grating.

strong illumination. Consider the direction θ such that the distances from the successive slits to the front of a plane wave traveling in this direction differ by one wave length. If a wave in this

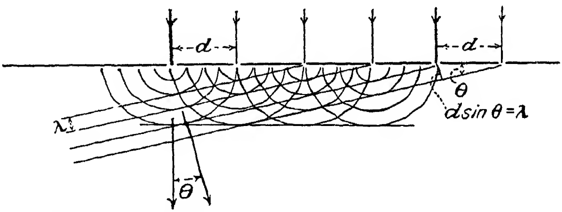


FIG. 277.—Detail of the action of a diffraction grating in terms of Huygens' principle.

direction is brought to a focus by a lens, all the different wavelets reach the same point in phase (strictly, differing in phase by 360°). They will therefore reinforce each other. The effect in terms of wavelets is shown for four slits in Fig. 277. This also shows the relation that must exist between the angle θ, the wave length λ, and d, the distance between the centers of adjacent

slits commonly known as the grating space. From the figure, we see that $\sin \theta = \lambda/d$. We can also have constructive interference when the path difference of the light from adjacent slits is twice the wave length or any whole number of wave lengths. The color of light depends on its wave length. If white light falls on a diffraction grating, the diffracted images of the source will come at different angles for different colors.

30. The gratings actually used in practice are almost always reflection gratings made by ruling lines on a polished metal surface with a diamond point. The diamond destroys the reflecting power of the metal so that only the spaces between rulings reflect the light. They then act as new sources exactly as in the transmission grating described above. Another trick used in most gratings is to rule the lines on a surface that is concave so that it automatically brings the light to a focus without the interposition of a lens. The grating space in gratings used to study visible light is very small, usually less than 1/10,000 in. The ruling of good diffraction gratings is so difficult that it has never passed the stage of a highly individualistic art. Professor Rowland at Johns Hopkins forty years ago made such good gratings that many of them are still in use all over the world. It is still impossible to buy a grating that is very good from an apparatus company. One must ask one of the two or three men in the world that make them for the privilege of purchasing one, and even then may have to wait years for a really good grating to be produced.

Spectra

31. When the light of a carbon arc or any other incandescent body is studied with a diffraction grating, it is found to consist of a continuous range of colors from deep red through all the hues of the rainbow to violet. This spread of colors formed by a diffraction grating or by refraction through a prism (or through the raindrops that produce the rainbow) is called a spectrum. In fact, the term is extended to include any record of the different wave lengths in a train of waves, visible or invisible. The wave lengths of the electromagnetic waves that are visible range from about $8{,}000 \times 10^{-10}$ m. for deep red to about $3{,}900 \times 10^{-10}$ m. for violet, although individual eyes vary considerably in their ability to see the ends of this range. The approximate relation of wave lengths and colors is shown in Fig. 278.

32. When the light of an electric discharge of a gas or vapor is studied with a diffraction grating, only certain definite colors are observed. Because the source of light used is ordinarily a line, the spectrum of such a discharge appears to be made up of a number of bright lines, each an image of the source in a different color corresponding to a different wave length. (Strictly speaking, there may be several wave lengths present which are easily separated into separate lines by the grating but which will differ so little that the eye cannot distinguish any difference in their color.) A spectrum of this sort is called a "line" spectrum. The number and arrangement of the "lines" in the spectrum, *i.e.*,

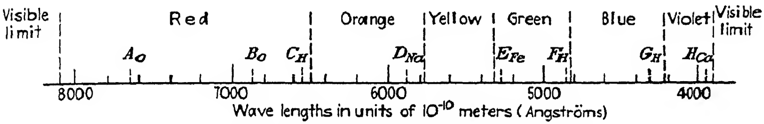


FIG. 278.—Wave lengths corresponding to the different colors. The dashed lines under letters are the Fraunhofer lines of the sun's spectrum. They are dark absorption lines appearing against the bright continuous spectrum. The subscript on the letter of each line is the symbol for the element causing the line. (*Cf.* Chapter XXVII, Par. 8.)

the wave lengths in the light given off by the source, are found to be characteristic of the gas or vapor in the source. It is found that no two substances have identical spectra when made luminous in the gaseous form. Sometimes if the source of light is a diatomic gas, the wave lengths present appear in groups that look like bands in instruments of low power. Such spectra are called band spectra.

33. The spectra we have been describing were those obtained directly from a luminous source and are called emission spectra. It is found that substances not only emit certain particular colors, but they absorb certain particular colors. Indeed, most of the color in the world around us is caused by unequal absorption and reflection of the various parts of the sun's spectrum by different substances. A piece of blue glass looks blue because it is opaque to all colors but blue, or, at least, the resultant psychological effect of the colors it does not absorb is blue. If we take a source that gives a continuous spectrum, a carbon arc for instance, send its light through a chamber full of the material whose absorption we want to study, and then spread the light out into a spectrum,

there will be black lines or bands in the spectrum corresponding to the wave lengths that are selectively absorbed by that particular material. Such a spectrum is called an absorption spectrum and, like an emission spectrum, is characteristic of the absorbing substance. The study of spectra has contributed enormously to our knowledge of atoms and molecules in the last fifty years.

X-Rays

34. Like radioactivity and electrons, x-rays were discovered in that brilliant last decade of the nineteenth century which laid the foundations for so much of modern physics. Röntgen in working with an electrical discharge in gases at low pressure (see Chap. XIV) found that radiations were given off by the tube which could cause fluorescence or affect a photographic plate and which would easily pass through screens that were opaque to ordinary light. He found that paper or wood or even thin sheets of aluminum let the rays through but that heavier substances, lead in particular, were much more effective in stopping them. He obtained photographic plates showing the shadows of the bones of the hand which evidently were more opaque to the rays than the tissues. These rays, which Röntgen christened x-rays, also had the property of discharging an electroscope. The student familiar with present-day medical and dental use of x-rays will recognize most of these properties, recorded by Röntgen in 1895, as ones that are of great practical value today. As to the origin of the rays, Röntgen reports: "The x-rays proceed from that spot where according to the data obtained by different investigators, the cathode rays strike the glass wall. If the cathode rays within the discharge apparatus are deflected by means of a magnet, it is observed that the rays proceed from another spot—namely, from that which is the new terminus of the cathode rays."

35. The study of x-rays has contributed tremendously to the development of science in ways that fall roughly into two classes, something like the uses of radioactivity: the first class comprises the study of the nature and origin of x-rays themselves, and the second class includes the innumerable uses of x-rays as tools, scientific, medical, and technical. Our treatment of both fields must be brief. We shall begin by describing the principal facts now known about the origin and nature of x-rays.

36. X-rays are now known to be electromagnetic radiation of very short wave length (10^{-11} to 10^{-9} m., approximately). This has been proved by showing that they are not deflected by magnetic or electric fields, that they undergo refraction and diffraction (although in so small a degree because of their small wave length that these effects escaped detection for a long time) and that they can be polarized. Their wave length has been measured by reflecting them from an ordinary diffraction grating at a grazing angle. Their approximate wave length was determined much earlier by studying their diffraction by crystals. The atoms in crystals are arranged in a definite pattern so that they form a natural three-dimensional diffraction grating of sufficiently small grating space to be suitable for the study of wave lengths in the x-ray region.

37. Röntgen's observation that x-rays originate when cathode rays strike the walls of the tube has been substantiated and a much greater intensity of x-rays obtained by concentrating the cathode rays in a discharge tube on a small metal target which then becomes the source of the x-rays. In early x-ray tubes, the cathode rays were obtained from the residual gas in an evacuated discharge tube. The thermionic effect is now used, and the cathode rays are the electrons emitted by a hot filament and accelerated by a strong electric field between the hot cathode and the target. The tube is evacuated as highly as possible so that the effects of residual gas are negligible.

38. The nature of the x-rays given off depends not only on the material of the target, but in a very interesting way on the voltage drop between the cathode and the target. This is the reason that all x-ray installations involve the transformers and high potentials that give a sinister and mysterious air to medical and dental installations. No matter how many electrons hit the target, no x-rays are produced unless there is a high potential across the tube. Specifically, it is found that the highest frequency ν (shortest wave length) given off by an x-ray tube is given by

$$\nu = \frac{Ve}{h} \tag{6}$$

where V is the voltage across the tube, e is the electronic charge, and h is a universal constant known as Planck's constant and

having the value 6.55×10^{-34} joule-sec.* (We shall have much more to say about this constant in the next chapter.) We believe that the crucial point about this equation is not the potential drop but the kinetic energy it gives to the electrons from the

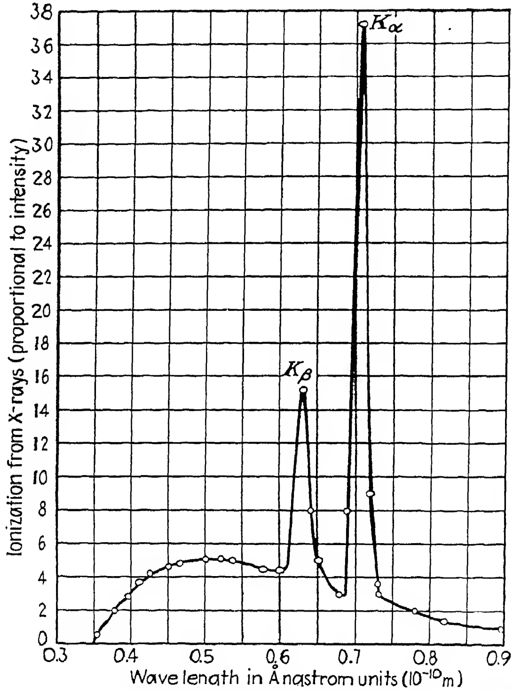


FIG. 279.—Part of the x-ray spectrum of molybdenum. (*After Ulrey.*)

filament. If they could be speeded up in some other way, they would produce the same effect. Therefore, we really should write Eq. (6) as

$$\nu = \frac{mv^2}{2h} \tag{7}$$

where m is the mass of an electron and v its velocity. There is no precise way of saying how short the wave length of an electromagnetic wave has to be for it to be an x-ray, but we can say

* 6.61×10^{-34} is probably more accurate.

roughly that if it is more than 10×10^{-10} m. it has little penetrating power and is useless for most x-ray purposes. In fact, the penetrating power of x-rays is inversely proportional to the cube of their wave length. The x-ray tubes in ordinary use for medical photography operate at voltages of 60 to 100 kv., although much higher potentials are being tried for therapeutic purposes.

39. Although the short wave length limit of the x-ray spectrum given out by a tube is determined by the voltage, the general character of the spectrum above that limit depends on both the voltage and the material of the target. In general, it consists of a continuous spectrum of all wave lengths on which is superposed a spectrum of a few strong lines. A typical example is given in Fig. 279. This line spectrum is the part that is characteristic of the target. In general, it consists of two or three groups or series of lines called the K, L, and M series in order of increasing wave length and decreasing penetrating power. The wave lengths of the characteristic lines increase regularly as targets are made of lighter and lighter materials. The first conclusive experiments on this effect were reported by Moseley in 1912. His photographs of the K series from calcium to copper are shown in Fig. 280. He showed that the square root of the frequency of either of the strong lines in this series is nearly proportional to the atomic number of the target. More precisely,

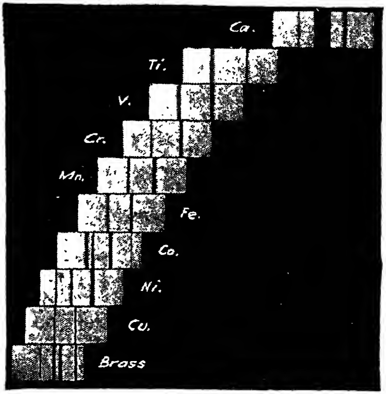


FIG. 280.—X-ray spectra (K series) taken by Moseley using various targets. The atomic numbers diminish steadily by one unit from copper, 29, to titanium, 22. Scandium is then missing and calcium has the atomic number 20. The wave lengths are increasing from left to right.

$$\sqrt{\nu} = A(Z - s) \qquad (8)$$

where Z is the atomic number and A and s are universal constants having the same value for all elements but differing values for different series and different lines in the series, *e.g.*, for the short-

est line in the K series, A and s would have the same value for all elements. This relation expressed by Eq. (8) is much simpler than exists between the visible spectra of the elements. It was one of the first discoveries that showed the atomic number to be of more fundamental significance than the atomic weight.

Application of X-Rays to Crystal Structure

40. We cannot take the time to discuss the innumerable medical and technical applications of x-rays, but there is one very important scientific application that we shall explain briefly. We have

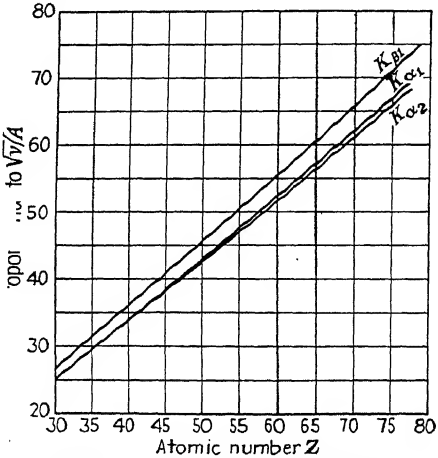


FIG. 281.—Moseley diagrams for three lines in the K series of x-rays for the elements from copper (Z = 29) to platinum (Z = 78).

already mentioned the diffraction of x-rays by crystals. In the case of certain simple crystals such as sodium chloride, the arrangement of the atoms of the crystals can be pretty well established by the methods of mineralogy. Furthermore, by knowing the density of the crystal and, from Avogadro's number, the number of atoms per gram, it is possible to calculate the spacing of the atoms in the crystal and therefore the grating space of the diffraction grating they form. From these data the first measurements of x-ray wave length were made. Once these wave lengths were measured, the diffraction patterns of crystals whose internal arrangement was unknown could be studied and from

these patterns the arrangement and exact spacing of the atoms in the crystal determined. In this way, a vast body of knowledge has been built up about the nature of metals, alloys, and nonmetallic compounds. It is a highly specialized subject that belongs as much in the field of chemistry or metallurgy as in physics.

Avogadro's Number from X-Rays

41. The original determinations of x-ray wave lengths came from the crystal measurements by a calculation involving Avogadro's number. The measurement of the wave lengths with ordinary optical gratings made it possible to reverse this process and use the crystal x-ray measurements to determine Avogadro's number and then to combine this value with the electrolysis results to get the electronic charge. This gave a value differing somewhat from Millikan's original value. The discrepancy has now been explained by new measurements of the viscosity of air which alter the value of the electronic charge deduced from Millikan's data. The new value which we quoted in Chap. XII (1.60×10^{-19} coulomb) is consistent with the x-ray optical-grating determination.

Gamma Rays

42. Of still shorter wave length than the x-rays are some of the gamma rays given off by natural or artificial radioactive substances. These are the electromagnetic waves sent out when the electrical structure of the atomic nucleus rearranges itself. The longer wave length gamma rays are of about the same wave length as the shortest of the x-rays, but the shortest gamma rays have wave lengths of as little as 10^{-13} m. It is possible that there is some electromagnetic radiation of about this wave length or less in the cosmic rays that come into the earth's atmosphere from space, but the exact nature of these rays is still doubtful.

43. We have now mentioned various categories of electromagnetic waves ranging in wave length from thousands of kilometers to the minute fraction of a centimeter given above. Until a few years ago, there were large gaps in this spectrum. Now these have all been closed. The shortest waves that can be produced by electrical circuits are shorter than the longest heat waves. The shortest heat waves merge indistinguishably with the visible

spectrum. The shortest waves produced in the ultraviolet by optical methods are shorter than the longest x-rays, and finally as we have seen the x-ray and gamma ray part of the spectrum overlap.

SUMMARY

The velocity of light was first measured by Roemer, who determined the time required for light from one of the moons of Jupiter to pass across the earth's orbit. In a recent determination by Michelson, light was reflected from one face of a rotating mirror to a distant plane mirror which returned it to another face of the rotating mirror. By measuring the final position of the light image, the speed of rotation of the mirror, and the various distances, the velocity of light was found to be 299,796 km./sec.

The most familiar properties of light are straight-line propagation, the law of reflection, and the law of refraction. For a reflected beam of light, the angle of reflection is always equal to the angle of incidence; for refracted light, the ratio of the sine of the angle of refraction to the sine of the angle of incidence is a constant called the index of refraction. The subject of geometrical optics is based on these laws.

The familiar properties of light are reconcilable with the wave nature of light if the wave length is very small. By using Huygens' principle that each point in the front of an advancing wave can be considered as a new center of disturbance, detailed explanations of refraction and reflection are worked out in terms of wave theory.

If light from a single bright source passes through two narrow slits close together and then falls on a screen, alternate bright and dark colored bands are seen. These are explained by the difference of phase between the light waves from the two slits. Newton's rings are another example of these "interference" phenomena. Similar effects are observed when light passes through a narrow slit. Such interference and diffraction phenomena are easily explained in terms of a wave theory of light but have never been satisfactorily explained in any other way. The diffraction grating is a device which uses this kind of phenomenon for the study of spectra.

The light given off by any incandescent substance contains a continuous range of wave lengths, *i.e.*, has a continuous spectrum.

Light from a gas or vapor through which an electric discharge is passing contains only certain definite wave lengths, *i.e.*, gives a line or band spectrum. If a continuous spectrum is passed through a gas or vapor, certain definite wave lengths are absorbed. Such wave lengths constitute an "absorption spectrum." The wave length of visible light varies between about $8{,}000 \times 10^{-10}$ m. for red to about $3{,}900 \times 10^{-10}$ m. for violet.

If high-speed cathode rays strike a solid target, x-rays are given off. X-rays, like visible light, are electromagnetic waves, but their wave length is so short that their properties are quite different from those of light rays. They are extremely penetrating, ionize the air through which they pass, and have dangerous physiological effects. The wave lengths of x-rays depend on the atomic number of the material in the target from which they come and on the energy of the electrons striking the target. If the voltage through which these electrons have fallen is V and their charge e, the highest frequency of x-rays given off is $\nu = Ve/h$ where h is a universal constant known as Planck's constant. Gamma rays from radioactive substances are electromagnetic waves of wave length ranging from that of short x-rays (about 10^{-11} m.) to as little as 10^{-13}.

ILLUSTRATIVE PROBLEMS

1. A beam of light traveling in glass strikes a plane glass-water boundary at an angle of 45°. What is the angle of refraction?

The index of refraction of water with respect to glass, Fig. 282, is defined as

$$n = \frac{\sin i}{\sin r} = \frac{v}{v'} \tag{9}$$

from Eq. (1), page 561, and Eq. (2), page 566. If we now define the index of refraction of a single medium as $n = c/v$ where c is the velocity of light in a vacuum, we may write Eq. (9)

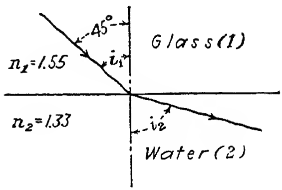


FIG. 282.—Refraction of light at a surface between glass and water. Illust. Prob. 1.

$$\frac{v_1}{v_2} = \frac{c/v_2}{c/v_1} = \frac{n_2}{n_1} = \frac{\sin i_1}{\sin i_2} = n$$

or

$$n_1 \sin i_1 = n_2 \sin i_2.$$

Here we have written i_1 for i and i_2 for r so that the equation holds no matter in which medium the light is incident. Now in Fig. 282, let glass with index of refraction $n_1 = 1.55$ be the first medium, and water with index of refraction $n_2 = 1.33$ be the second. Then

$$1.55 \sin 45° = 1.33 \sin i_2$$
$$\sin i_2 = \frac{1.55}{1.33}\frac{1}{\sqrt{2}} = 0.824$$
$$i_2 = \arcsin 0.824 = 55°29'.$$

i_2 in this problem is the angle of refraction.

2. The distance to the screen in Young's experiment is 2 m. The distance from the central bright spot to the third dark band is 5 mm. If the cadmium red line of wave length $6{,}438 \times 10^{-10}$ m. is used, how far apart are the two pinholes?

From Eq. (4), page 568, for the displacements of the dark bands, we have

$$x_n = \frac{L}{d}\left(\frac{2n-1}{2}\right)\lambda \qquad \text{where} \qquad n = 1, 2, 3, \cdots$$

so that for the third dark band $n = 3$, and

$$x_n = \frac{L}{d}\frac{5}{2}\lambda \qquad \text{or} \qquad d = \frac{L}{x_n}\frac{5\lambda}{2} \tag{10}$$

where d = the distance between the pinholes.

$x_n = 5 \times 10^{-3}$ m. = the distance of the third band from the central bright spot.

$L = 2$ m. = the distance from the pinholes to the screen where the bands are observed.

$\lambda = 6{,}438 \times 10^{-10}$ m. = the wave length of the light used.

With these values in Eq. (10), we find

$$d = \frac{2\text{ m.} \times 5 \times 6{,}438 \times 10^{-10}\text{ m.}}{5 \times 10^{-3}\text{ m.} \times 2} = 6{,}438 \times 10^{-7}\text{ m.}$$
$$= 0.6438\text{ mm.}$$

3. A plane diffraction grating has 5,000 lines per centimeter. The light from a mercury arc is incident perpendicularly on the grating. What will be the angular separation in the second order of the green line of $5{,}461 \times 10^{-10}$ m. and the line of $5{,}791 \times 10^{-10}$ m. wave length?

From Fig. 277, if the path difference is twice the wave length, we find that

$$\sin\theta = \frac{2\lambda}{d}.$$

A path difference of two wave lengths gives what is called the second-order spectrum. Here d is the distance between the lines on the grating and is

equal to 1/5,000 cm. $= 2 \times 10^{-6}$ m. For the line of $5{,}461 \times 10^{-10}$ m. wave length,

$$\sin \theta = \frac{2 \times 5{,}461 \times 10^{-10}\text{ m.}}{2 \times 10^{-6}\text{ m.}} = 0.5461$$

$$\theta = \arcsin 0.5461 = 33°6'.$$

For the line of $5{,}791 \times 10^{-10}$ m. wave length,

$$\sin \theta' = \frac{2 \times 5{,}791 \times 10^{-10}\text{ m.}}{2 \times 10^{-6}\text{ m.}} = 0.5791$$

$$\theta' = \arcsin 0.5791 = 35°23'.$$

The angular separation is therefore

$$\theta' - \theta = 35°23' - 33°6' = 2°17'.$$

PROBLEMS

1. How many revolutions per second must the mirror make in Michelson's measurement of the velocity of light?

2. Show that, if the speed of rotation of Michelson's mirror was determined to 1 part in 10^5, the time can be calculated to within 2×10^{-9} sec.

3. A ship at sea sends signals ashore by a light flash and underwater sound. If the sound traveling 1,450 m./sec. arrives 4 sec. after the light, how much difference does the finite velocity of light make in calculating the distance from the ship to shore?

4. A beam of light passing through water strikes a piece of glass at an angle of incidence of 30°. What is the angle of refraction?

5. Show that in passing from a more dense to a less dense medium the sine of the critical angle is equal to the index of refraction. The critical angle is the angle of incidence for which the refracted ray is in the plane separating the mediums.

6. Show that the positions of the bright bands in Young's experiment are given by

$$x_n = n\frac{L}{d}\lambda \qquad n = 0, 1, 2 \cdots.$$

7. The distance to the screen in Young's experiment is 5×10^3 times the distance between the pinholes. The average distance between dark bands is 2.95 mm. Find the wave length of the light used.

8. Show that the difference in path between rays reflected at the front and back surface of a thin layer of thickness d and index of refraction n is $2dn \cos r$ where r is the angle of refraction.

9. Let n represent the number of wave lengths difference in phase in the diffraction grating. It is known as the order of the corresponding spectrum. If a sodium line is diffracted at 60° in the third order, where will the first order fall?

10. What is the shortest x-ray wave length that can be produced by million-volt electrons?

11. How fast must the electrons in an x-ray tube travel to produce x-rays with a frequency of 3×10^{20} vibrations per second?

12. The wave lengths of the K_{β_1} series from calcium to copper are given in the following table.

Element	Atomic No.	K_{β^1}
Ca	20	3.35169×10^{-10} m.
Sc	21	3.02503
Ti	22	2.74317
V	23	2.49835
Cr	24	2.28503
Mn	25	2.09751
Fe	26	1.93208
Co	27	1.78529
Ni	28	1.65450
Cu	29	1.53726

Evaluate the constants in the equation representing this series from the first and last wave lengths, and calculate the other wave lengths.

13. Calculate the spacing between successive planes of atoms in the simple cubic crystal of sodium chloride.

14. The potential drop across an x-ray tube is 30,000 volts and the current flowing through it 10 milliamp. Assuming that 98 per cent of the energy acquired by the electrons between the cathode and the anode goes into heating the anode, how large a current of water must flow through the anode to keep it at constant temperature if the temperature of the water is raised 10° by going through?

CHAPTER XXVII

RADIATION AND MATTER

1. The last chapter has given a brief account of the overwhelming mass of evidence for the wave nature of light which has accumulated since 1801, when Young discovered the phenomenon of interference between the beams of light coming from two slits. During the nineteenth century, success followed success in the application of the wave theory of light to the many complicated phenomena of diffraction and interference. The work of Maxwell, interpreting light as an electromagnetic radiation, bound together in one theory experimental results from widely scattered fields. At the beginning of the present century, physicists were in practically complete agreement that the last word regarding the nature of electromagnetic radiation had been said and that further developments of the theory of light would be in the direction of amplifying and extending the theory along lines which had already been laid down by Maxwell.

This happy state of satisfaction has since been destroyed by a number of experiments which have necessitated a return to a modified form of the corpuscular theory of light. It is our purpose in the present chapter to outline the experimental results that led to the breakdown of the simple wave theory of light and to describe the current ideas concerning the nature of radiation. At the same time, we shall give a brief discussion of atomic and nuclear structure.

Where Does Light Come From?

2. In the first place, we must point out explicitly that in the electromagnetic theory of light the source of the electromagnetic radiation lies in *accelerated* charged particles. Thus, radio waves are produced by electrons rushing back and forth in the antenna of a transmitting station. At the target of an x-ray tube, high-speed electrons strike the dense material of the target and are

stopped. The kinetic energy which the electrons possessed goes chiefly into heating of the target, but a small part of it flies away with the speed of light in the x-radiation which we have described as an electromagnetic radiation of short wave length; the deceleration of the electrons in the target has been accompanied by the radiation of energy in the form of x-rays.

3. Visible light has its source in "excited" atoms. You will remember that the current picture of the structure of an atom places a small nucleus at the center of the atom, having a positive charge equal to Z electronic charges, where Z is the atomic number of the atom in question. Practically the entire mass of the atom is concentrated in the nucleus. Around the nucleus rotate Z electrons which are located at various distances from the nucleus, the orbit of the outermost electron defining in a rough way the diameter of the atom as a whole. Now it is reasonable to inquire why, if an accelerated electron is capable of radiating energy, the single negative electron circling about the nucleus of an atom of ordinary hydrogen does not slow down, radiating away its kinetic energy as light, and fall into the nucleus, which attracts it by virtue of its opposite electric charge. Just as the centrifugal force of the moon's rotation about the earth counterbalances the gravitational attraction between moon and earth, and keeps the moon from falling on us, so the centrifugal force of the electron in its orbit balances the electrical attraction of the nucleus for the electron. But the moon has no way of getting rid of its kinetic energy, whereas the electron is provided with the possibility of shedding its kinetic energy as radiation; why, then, does it not do so?

4. The answer to this question involves so much that we have not mentioned that we shall not attempt to give it here but shall content ourselves with describing the conditions found necessary for the success of our atom model. One such condition is that only certain electron orbits are possible, and another condition is that each orbit contains only one electron. Since the positively charged nucleus exerts an attraction on the extranuclear electrons, work must be done on an electron to remove it entirely from the atom (*i.e.*, to ionize the atom), and this work will be greater the closer the electron originally was to the nucleus. To each electronic orbit in a particular atom, then, corresponds a certain ionization energy, which is just the work necessary to

remove an electron from that orbit and take it to a great enough distance so that it is completely out of the influence of the nucleus. No two permitted electronic orbits in an atom will have just the same ionization energy, and we may speak of the energy difference between them as being the difference of the ionization energies of the two. An atom left to itself will assume the state in which it is most stable, *i.e.*, the state in which all its extranuclear electrons are in orbits as close to the nucleus as possible, consistent with the restrictions stated above that only certain orbits are possible and that each of these can contain only one electron.

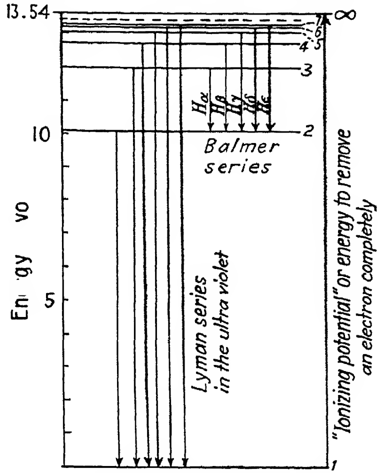


FIG. 283.—Stationary states in a hydrogen atom. The single electron revolves only in certain permitted orbits about the single proton which forms the nucleus. Normally it is in the innermost orbit. The energies needed to move it to other permissible orbits, or given up by it on return from them, and the corresponding spectra, are indicated here.

Emission Spectra

5. This model of the atom may seem a very complicated and artificial one, but it is required to explain all of the features of spectra of the elements, spectra being themselves very complicated. It is found experimentally that a gas which has been excited to emit light (as the gas in a neon sign, for example) sends out not a continuous spectrum containing all colors, but a set of sharply defined wave lengths which are characteristic of the particular gas which emits them. In terms of our atom model, these wave lengths are interpreted in the following way: swift electrons traveling through the gas because of the electrical discharge strike atoms in their path and excite them; the electrons give up some of their kinetic energy to the atoms. Now the only way in which energy can be absorbed into the electronic structure of an atom is in shifting an electron from its stable orbit near the nucleus to another permitted orbit farther away. In particular, the new orbit may be infinitely far from the nucleus, *i.e.*, the atom

may be ionized by electron impact. But the important thing to notice about this process is that the energy absorbed from the electron by the atom must have one of a set of discrete values, namely, a value that corresponds to the energy difference between two stationary states of the atom—two states in which all of the electrons are rotating in permitted orbits about the nucleus. The second stationary state of the atom will be an excited state, in which one of the electrons is not in the lowest orbit permitted it, and the electron very soon spontaneously jumps down from its new orbit to the old one it occupied before the electron impact, radiating as light the difference in energy between the states. The frequency of the radiated light is determined by the energy

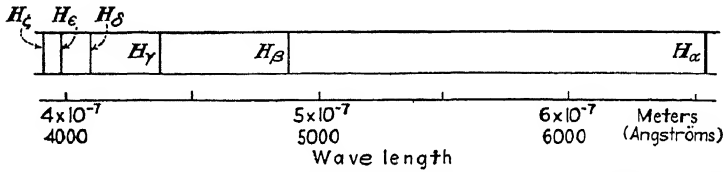


FIG. 284.—The spectrum of atomic hydrogen in the visible region. These lines constitute the Balmer series. Their regularity gave the first clue to the law governing the existence of stationary states. (*After a photograph.*)

which must be got rid of, *i.e.*, by the energy difference between the excited state and the normal state of the atom. This is the crux of the whole affair. If the energy difference between the normal state of an atom and the excited state in which it exists after an electron impact is given by E joules, then the frequency of the light which will be radiated when the atom returns to the normal state is given by $E = h\nu$ where h is Planck's constant which we have encountered already in the last chapter. Its value is 6.55×10^{-34} joule-sec.*

6. The spectrum of an excited gas, then, consists of bright lines, rather than a continuum, because of the discrete energy differences between the stationary states of the particular atom concerned, and the line spectrum is characteristic of the gas because each element has its particular arrangement of stationary states, the energy differences among them being different from those among the stationary states of any other sort of atom. This must be so, because the charge on the atomic nucleus deter-

*6.61×10^{-34} is probably more accurate.

mines the arrangement of permitted stationary states and the atoms of different elements are distinguished by differing nuclear charges.

7. The spectrum of an incandescent solid is a continuous one, but this is not in contradiction to the hypothesis made above, for in a solid the atoms are packed so closely to one another that the interaction between one atom and its neighbors cannot be neglected. The stationary states of a particular atom, then, will be influenced by the neighboring atoms, and the energy differences will be altered enormously. The spectrum of discrete lines which characterizes an element in the gaseous state is smeared out into a continuous spectrum when the element is a solid, and the resulting continuum conveys no information whatever about the chemical nature of the solid.

Absorption Spectra

8. The atoms of a gas which we considered in the last section were excited by electron impacts and subsequently went back to their most stable state with the emission of radiation. The inverse process also occurs; if light of just the proper frequency falls on an atom, an electron may be raised from its lowest stable orbit to another permitted one, and the energy necessary to do this is absorbed from the incident light. Subsequently, the atom returns to its original state, reemitting the radiation that it absorbed. This phenomenon may be observed, for example, in the spectrum of the sun as viewed from the surface of the earth. The sun is an incandescent solid, so that it emits a continuous spectrum, whereas the atmosphere of the sun is an absorbing gas made up of the elements composing the sun in gaseous form. It was observed long ago that the spectrum of the sun consists of dark lines on a bright background, *i.e.*, is just the inverse of the bright-line spectrum of an excited gas. In terms of the atom model that we have outlined, the dark lines of the solar spectrum represent the frequencies that correspond to the energy differences between pairs of the stationary states of the atoms forming the solar atmosphere. They correspond exactly in wave length to bright lines in the emission spectra of these elements, and so it has been possible by a study of the solar spectrum to identify the elements present in the sun. (*Cf.* Fig. 278. page 573.)

The Photoelectric Effect

9. All that has been said so far in this chapter can be accounted for on the basis of the electromagnetic theory of light, if one is willing to stretch the theory far enough. The greatest wrench to the theory has come in the assumption of stationary states of atoms, in which the electrons are pursuing circular or elliptical orbits about the nucleus; for uniform circular motion must be accompanied by constant central acceleration (Chap. XVIII), and the electromagnetic theory states than an accelerated electron *must* radiate. However, if one is willing to say that certain orbits are possible in which the electrons can rotate stably without radiating, whereas radiation must take place if the electron leaves such an orbit, and must continue until it has reached the next lower stable orbit, then we may still preserve the electromagnetic theory.

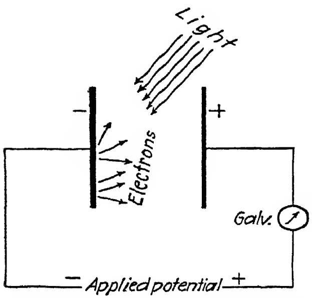


FIG. 285.—Schematic diagram of the photoelectric effect.

10. We have now to consider an experimental phenomenon for which no explanation on the basis of the electromagnetic wave theory of light is possible; this is the photoelectric effect. The experimental data are as follows: When light falls on a metal plate, electrons are sometimes emitted from the surface of the plate, and it is found that no electrons are emitted from the plate until the frequency of the light falling on the plate has reached a certain well-defined lower limit, *regardless of the intensity of the incident light.* Moreover, once the light falling on the plate is of short enough wave length (*i.e.*, of high enough frequency) to cause the emission of electrons from the metal, any increase in the intensity of the light causes an increase in the *number* of the emitted electrons, *but the energy of each electron is fixed by the frequency of the light and can be increased only by increasing the frequency.* The experimental data regarding this effect can be summed up in the simple expression proposed by Einstein in 1905:

$$\tfrac{1}{2}mv^2 = h\nu - \phi \tag{1}$$

where $\frac{1}{2}mv^2$ is the kinetic energy of the ejected photoelectron, h is Planck's constant, ν the frequency of the incident light, and

ϕ a quantity called the work function that differs from metal to metal and expresses the energy in joules which must be expended to remove the electron from the metal. Extremely careful experiments by Millikan and others showed that this expression accurately described the experimental results.

11. Now there are two features of this experiment that are in violent disagreement with predictions of the electromagnetic wave theory of light. The first is simply that in that theory no dependence of the energy of the emitted electron on the frequency of the incident light is to be expected, whereas such a dependence is unmistakably present in the experimental data. The second difficulty has proved to be completely insurmountable and depends on the fact that the electromagnetic wave theory of light allows one to calculate the rate at which energy is being carried from the radiating to the absorbing body in a beam of light of measured intensity. If one calculates, from the theory, the number of joules per second that are being carried by the light of a distant star to a small metal plate in the laboratory, it may turn out that a considerable fraction of a second must elapse before enough energy has been communicated to the plate by the light to permit the emission of one photoelectron. Now the question of how the plate stores up this energy for, say, a hundredth of a second and then delivers it *all* to one photoelectron is bad enough, but this is not the worst. For one can find experimentally that although, *on the average*, the time between the illumination of the plate and the ejection of the first photoelectron is that predicted by the theory—perhaps a hundredth of a second —*sometimes* the first photoelectron will appear only a thousandth of a second after the plate is illuminated. Consider this situation: Before the light has had time to transfer a total amount of energy E to the plate (according to the wave theory), a photoelectron of energy $10E$ has been knocked out of the plate. It is apparent that a good deal of stretching and twisting of the wave theory would be necessary to make it cover this phenomenon. As a matter of fact, it is clear that this phenomenon is best interpreted by a modified corpuscular view of light, in which a beam of light is not viewed as a wave motion at all, but as a stream of particles, each of energy $h\nu$, where ν is the frequency of the light. These particles, regardless of their energy (or frequency) travel with a fixed velocity: the velocity of light.

The Compton Effect

12. In 1923, A. H. Compton discovered a phenomenon which bears his name and which provides further evidence that it is sometimes necessary to conceive of light as consisting of particles. Briefly put, the Compton effect is the following: When x-rays of initially homogeneous wave length (*i.e.*, all initially of the same wave length) fall on matter, the original beam is scattered in all directions *and at the same time the wave length is changed, the change in wave length depending on the direction of scattering, and only on the direction of scattering.* Suppose, for example, that x-rays of wave length 100 X units (1 X.U. $= 10^{-13}$ m.) fall on a carbon scatterer. If the radiation coming from the

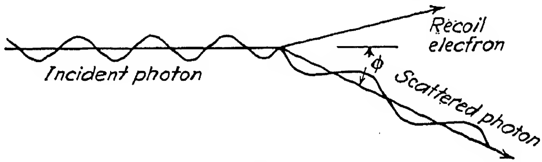


Fig. 286.—Schematic diagram of the Compton effect.

carbon at an angle of 90° with the initial direction of the x-ray beam is examined, it will be found to consist of x-rays of wave length 124 X.U. The radiation scattered backward from the carbon toward the source (angle of scattering 180°) will have a wave length of 148 X.U. To this change in wave length corresponds a change in frequency, and hence a change in what we have spoken of above as the energy of the radiation, which is related to the frequency by the relation $E = h\nu$. The results described above with a carbon scatterer are unchanged, whatever element is employed as a scatterer, so that the entity responsible for the scattering must be one which is present in all atoms. The simplest hypothesis that suggests itself is that the extranuclear electrons are responsible for the scattering, since all atoms possess these. Although it has been impossible to explain the Compton effect in terms of the wave theory of light, a complete and simple explanation was given by Compton on the basis of the particle view of radiation. If a beam of light is regarded as being made up of particles, which we may call *photons*, each of energy $h\nu$, where ν is the frequency of the light in terms of the wave theory, then one can conceive of an elastic collision between such a photon

and an electron. In such a collision, some of the energy of the photon is transferred to the electron, which moves off with its newly acquired kinetic energy, while the photon is deflected through some angle φ, with reduced energy. Since the frequency of the light is directly proportional to its energy, the frequency must have been lowered by the collisions, *i.e.*, the wave length of the radiation has been increased by the scattering. The momen-

FIG. 287.—Track of an electron produced by Compton scattering and deflected by a magnetic field perpendicular to the plane of the photograph. (*Delsasso, Fowler, and Lauritsen.*)

tum of a photon is its energy divided by its velocity, namely, $h\nu/c$, where c is the velocity of light. If the conditions of the collision are those shown schematically in Fig. 286, the conservation of energy requires that

$$h\nu = \tfrac{1}{2}mv^2 + h\nu'. \tag{2}$$

The conservation of momentum in the horizontal direction requires that

$$\frac{h\nu}{c} = mv \cos\theta + \frac{h\nu'}{c}\cos\varphi, \tag{3}$$

and conservation of momentum in the vertical direction requires that

$$0 = mv \sin \theta - \frac{h\nu'}{c} \sin \varphi. \tag{4}$$

13. These three equations* taken together enable one to solve for the unknowns: the velocity v of the electron after the collision, the angle between the direction of the ejected electron and the incident photon, and the new frequency ν' of the photon after the collision, for any particular angle φ in which the scattered radiation is observed. For example, the new wave length of light which has been scattered at an angle φ with its original direction is given by

$$\lambda' = \lambda + \frac{h}{mc}(1 - \cos \varphi) \tag{5}$$

where λ is the original wave length, m the electronic mass, c the velocity of light, and h Planck's constant.

Waves or Particles

14. The whole of modern physics has been colored by the following dilemma, which is very sketchily indicated in the last chapter and in this one: In experiments involving diffraction and interference of light, a theory which treats light as a wave motion is completely successful; yet in the fundamental processes whereby light interacts with electrons, such as the photoelectric effect and the Compton effect, the experimental results simply cannot be explained on the basis of the wave theory, but demand

* If we take account of relativity corrections, Eqs. (2), (3), and (4) should read

$$h\nu = m_0c^2\left(\frac{1}{\sqrt{1-\beta^2}} - 1\right) + h\nu' \tag{2'}$$

$$\frac{h\nu}{c} = \frac{m_0v}{\sqrt{1-\beta^2}} \cos \theta + \frac{h\nu'}{c} \cos \varphi \tag{3'}$$

$$0 = \frac{m_0v}{\sqrt{1-\beta^2}} \sin \theta - \frac{h\nu'}{c} \sin \varphi. \tag{4'}$$

where $\beta = v/c$.

The solution of the equations in this form leads directly to (5). Equations (2), (3), and (4) lead to (5) only approximately.

that light be regarded as consisting of particles of energy $h\nu$. The fact that the particle view of light cannot escape completely from the bounds of the wave theory is shown by the way in which the words "wave length" and "frequency" keep occurring in our discussions of photons. Our theory of light as a wave motion needs revision, it is apparent, but the revision must be such that

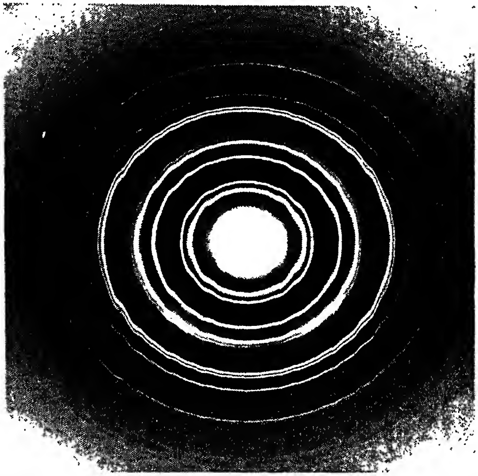


FIG. 288.—The diffraction of electrons by passage through aluminum 2×10^{-8}. m. thick. This pattern is similar to that obtained by the diffraction of x-rays under corresponding circumstances. The electrons in this photograph had an equivalent wave length of .056 A.U. or 5.6×10^{-12} m (*Courtesy of Dr. L. H. Germer.*)

the successes of the wave theory in explaining experimental results are preserved.

The Diffraction of Material Particles

15. A clue to the reconciliation of the wave-particle controversy is to be found in the breakdown of the classical concept of a particle when this concept is applied to very small particles indeed, namely, the electron, the proton, or the atoms of the lightest elements. In 1929, Davisson and Germer found that, when a beam of electrons of uniform energy was directed onto a crystal of nickel, the electrons were reflected most strongly in certain directions, and when the energy of the electrons was changed, the

directions of selective reflection of the electrons also changed. Now this is a phenomenon suspiciously like the reflection of light from a diffraction grating, where the lines of the grating are here replaced by the lines of nickel atoms in their orderly array in the crystal. It was found, in fact, that this experiment and others of the same sort could be completely explained if the beam of electrons were regarded as a "wave motion" like light or x-rays, of wave length h/mv, where h is the ubiquitous Planck constant, m the mass of each electron, and v the velocity of the electrons. The only way in which the experiment of Davisson and Germer could be explained was to assume that it was a true interference phenomenon.

16. Similar diffraction and interference effects have been observed even for hydrogen molecules, although the masses of these particles are so much greater than the electronic mass that their wave lengths are very short unless they are traveling very slowly indeed.

17. In the last few years, then, experimental evidence has been adduced to show the particle-like nature of radiation, which had been regarded as a wave motion, and the wavelike nature of entities, which had always been regarded as particles. At present, it seems a fair generalization from our knowledge to say that the concepts "wave" and "particle" lose their distinctness when they are applied to systems of atomic dimensions. The concept of a particle is that of an object differing from a billiard ball only in size, a body that obeys the laws of mechanics which we apply with such success to billiard balls themselves. Now what is suggested by the experiments that have been discussed in this chapter is that the laws of mechanics which seem to describe so well the motion of billiard balls are simply extremely good approximations to the truer laws of mechanics which we have found to apply to electrons. These latter laws must include the phenomena of diffraction and interference, for such phenomena are observed in the motion of electrons. They are, in fact, the laws of the "wave mechanics," which has had such conspicuous success, since its initiation a decade ago, in describing the motions and interactions of electrons, atomic systems, and radiation. Although the mathematical methods of the wave mechanics are still so unfamiliar as to seem extremely difficult, in effect the motions of particles are dealt with by considering the behavior of

waves whose wave length is given by h/mv, the intensity of the wave motion at a given point then expressing the likelihood that a particle will be found at that point. In just the same way, the whole machinery of the wave theory of light is retained in the modern theory of light, in so far as the propagation of the radiation is described in the same terms. Where the older theory of light interpreted the intensity of the wave motion at a given point as the intensity of light at that point, the modern theory takes the view that the intensity of the wave motion at a given point gives the probability that one will find a photon at that point—gives, in other words, the *average* number of photons arriving at that point each second. There is no guarantee, in the modern theory, that the photons will arrive uniformly, whereas the old theory did provide a uniform flow of energy. The whole point of this discussion is that neither the idea of a particle, which one gets by considering a baseball, nor the idea of a wave, which one gets by watching the ocean, is completely applicable to what are today called the fundamental entities of nature: electrons, protons, neutrons, and radiation. A fusion of the two ideas is demanded by an overwhelming array of experimental evidence.

18. Presumably, the laws that describe the motion of baseballs and billiard balls are those of the wave mechanics, but you will remember that phenomena of diffraction and interference are observable only when they are looked for in detail comparable with the wave length of the entity considered. The wave length of the regulation league baseball traveling 30 m./hr. is about 3.5×10^{-35} m., so that it is clear that deviations from classical mechanics in its motion are completely unobservable. It is, of course, for this reason that the wave mechanics was not invented until physics had to deal with the motions of "particles" unimaginably smaller than baseballs.

The Structure of the Nucleus

19. We have heretofore said very little about the way that the nuclei of atoms are made; they have remained for us heavy, small, positively charged bodies situated in the center of the atom. Sometimes they come apart spontaneously (Chap. XXI), but more often they stick together. We shall now consider briefly what is known about the constituents of nuclei and the way in which they are built. This branch of physics is one of the most

recent and rapidly progressing fields of investigation, and advances in our knowledge regarding nuclei are being made at an astonishing rate.

20. You will recall that a given nucleus is designated by its mass number and its charge number, or atomic number. The latter number determines the number of extranuclear electrons possessed by an atom and, hence, its chemical properties. The chemical elements are different from one another because of differing charges on their respective nuclei. Nuclei having the same charge need not all have the same mass, as we have already seen (Chap. XX), and, in general, an element will consist of a mixture of two or more *isotopes*, having the same charge number, but differing in mass number. Before proceeding to a more extensive discussion of nuclei, we must describe one of the constituents of almost all nuclei, the neutron.

The Neutron

21. The neutron is a particle of mass almost exactly that of the proton and of electrical charge zero. It exists in most nuclei, as has been inferred from the fact that neutrons are emitted when nuclei are disintegrated by the impact of swift bombarding particles, a process that will be considered later in this chapter. Neutrons produce practically no effect on electrons, so that they do not ionize atoms through which they pass by knocking electrons from these atoms; this makes it impossible to trace the path of a neutron in a cloud chamber. Neutrons do, however, quite often execute elastic collisions with atomic nuclei. In such a collision, the neutron imparts to the nucleus a fraction of its kinetic energy, and the struck nucleus becomes a swiftly moving charged particle, whose path may be studied in the cloud chamber. A study of the tracks of such "recoil nuclei" is one of the most fruitful ways of gaining information regarding the neutron.

Nuclear Binding

22. Probably the first question that suggests itself to the ingenious student is: What keeps nuclei together? It is manifest that, if nuclei are built of positively charged protons and uncharged neutrons, the only electrical forces within the nucleus can be those of repulsion, and since the nucleus is very small (about 10^{-14} to 10^{-15} m., from alpha-particle scattering experi-

ments) these repulsions will be very large indeed, since the protons are so close together. The electrostatic force between two protons 10^{-15} m. apart is $e^2/4\pi\epsilon_0 r^2$, or $\dfrac{(1.6 \times 10^{-19})^2 \times 9 \times 10^9}{10^{-30}}$, which comes out to be 230 newtons. This is an extremely big force, and one can readily see that it would produce a tremendous acceleration if either proton were free to move, since the mass of the proton is about 1.66×10^{-27} kg. We are compelled, then,

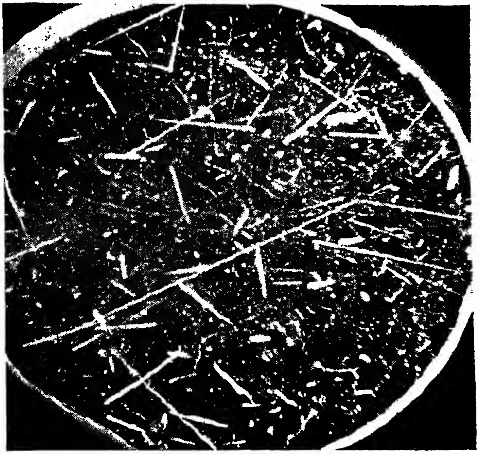


FIG. 289.—Tracks of atomic nuclei that have been struck by fast neutrons. The source of neutrons was in a cyclotron many feet away. (*Courtesy of Professor E. O. Lawrence.*)

to postulate some sort of attractive force among the particles which go to make up a nucleus, a force that must have enormous values when the particles are quite close together but decreasing so rapidly with increasing distance between particles that its effect vanishes outside a distance of the order of magnitude of the "size" of the nucleus: about 10^{-14} m. When nuclear particles (*e.g.*, a swift alpha particle and a nitrogen nucleus) do not come closer to one another than 10^{-14} m., the forces between them are wholly electrostatic in nature, a conclusion that is established by the complete experimental verification under the preceding conditions of the laws of alpha-particle scattering worked out by Rutherford entirely on the basis of Coulomb's law of force between charged particles. When nuclei approach one another

more closely than about 10^{-14} m., however, we may expect the short-range attractive force, which we have adduced to explain the fact that nuclei do not ordinarily explode of themselves, to begin to cancel, and finally to overcome the Coulomb repulsion, so that alpha-particle scattering by nuclei under conditions of extremely close approach should not obey the simple law of Rutherford. This prediction is experimentally verified. Deviations from the Rutherford scattering law were first found by Chadwick and Bieler in 1921, in a study of the scattering of alpha particles by hydrogen nuclei, and recently the existence of these short-range attractive forces between protons has been shown unmistakably.

Mass Defects of Nuclei

23. It has already been stated (Chap. XX, Pars. 19 and 20) that the masses of all atomic species are nearly, but not quite, whole numbers on the mass-number scale universally used: that in which the mass of O^{16} is exactly 16.0000.* The divergences from whole numbers indicate, since all nuclei are built of the same elementary particles, protons and neutrons, changes in the masses of these particles that depend on the sort of nucleus into which the particles are built. To illustrate: The mass of an atom of light hydrogen is about 1.0081, and the mass of a single free neutron is about 1.0090. Yet the mass of an atom of deuterium, or "heavy" hydrogen, is 2.0147, in spite of the fact that we consider its nucleus as being built of a proton and a neutron closely bound together by the intranuclear force already mentioned in Par. 22 of this chapter.

$$
\begin{array}{lcl}
\text{Mass of } {}_1H^1 & = & 1.0081 \\
\text{Mass of } {}_0n^1 & = & \underline{1.0090} \\
\text{Sum of } {}_1H^1 & & \\
\quad \text{and } {}_0n^1 & = & 2.0171 \\
\text{Mass of } {}_1H^2 & = & \underline{2.0147} \\
\text{Difference} & = & 0.0024
\end{array}
$$

* It should be explained here that, in what is to follow, it will be important to have a brief designation of each atomic species. We shall designate the charge number of an atom by means of a subscript to the left of the chemical symbol of the element of which the atom is an isotope and the mass-number by means of a superscript to the right. That is, the isotope of neon of mass number 20 will be denoted by ${}_{10}Ne^{20}$.

What has become of the difference in mass between H^1 plus a neutron, and H^2? The discussion at the end of Chap. XX of the equivalence between mass and energy gives us the clue. Just as a particle becomes heavier when work is done on it to accelerate it, so it will become lighter if it gives up energy. It is apparent that a proton and a neutron bound tightly to one another by our intranuclear force require work to pull them apart; *i.e.*, energy must be added to a deuterium nucleus to separate it into an isolated proton and an isolated neutron. By the equation relating mass and energy, we are enabled to guess that the energy required to separate a deuteron (as the nucleus of H^2 is called) into a proton and a neutron will be given in joules by c^2 times the difference between the deuteron mass and the sum of the masses of proton and neutron, if this mass difference is expressed in kilograms. A free proton and a free neutron coming together to form a deuteron would give off precisely the same amount of energy, which would be radiated at the expense of their total mass. We see, then, that the small divergences of atomic masses from whole numbers afford a clue to the *binding energies* of nuclei, *i.e.*, to the energies necessary to break the nuclei up into their component parts. It is customary, instead of expressing energies in joules and masses in kilograms, to use for the purposes of nuclear physics a million electron volts (mev) as the unit of energy, and $\frac{1}{16}$ of the mass of O^{16} as the unit of mass. The relations between them are: 1 mev $= 1.074 \times 10^{-3}$ mass units (m. u.), and 1 mass unit $= 930.95$ mev, 1 mev $= 1.60 \times 10^{-13}$ joule.

Formation of Electrons and Positrons

24. Beta particles, which we have discussed in the chapter on radioactivity, are electrons emitted from the nucleus. In spite of this fact, there has been nothing at all said in the present chapter about the presence of electrons in the nucleus. This is so because there are compelling theoretical reasons for believing that electrons cannot exist in the nucleus. The emission of them in radioactive changes is explained on the hypothesis that the proton and the neutron can change into one another: that the neutron can, given enough energy, transform itself into a proton; this change being accompanied by the emission of an electron which has at that instant been created from an amount of energy necessary to form its mass in accordance with the equation

$E = mc^2$. The energy required to create an electron is about 0.5 mev. Similarly, the proton is regarded as being capable of transforming itself into a neutron, by the creation and emission of a *positron*, provided sufficient energy is available. The positron, a particle we have not met before, was discovered in 1932 by Dr. Carl Anderson, at the California Institute of Technology. It has a mass equal to that of the electron and a positive charge

FIG. 290.—"Pair production." Tracks of a positron and an electron produced simultaneously in air by a gamma ray. A magnetic field perpendicular to the plane of the photograph has curved the tracks in opposite directions because of their charges of opposite sign. (*Delsasso, Fowler, and Lauritsen.*)

which is equal in magnitude, but opposite in sign, to that of the electron. Positrons exhibit the remarkable phenomenon that when a positron meets an electron sufficiently closely (and it is obvious that they will attract one another) both particles disappear, and the energy associated with their mass flies away with the speed of light as gamma radiation. This process is known as *annihilation* of positrons and is imperfectly understood. Even more remarkable is the process known as "pair production" in which an electron and a positron are created simultaneously from a photon of gamma rays (Figs. 290 and 293).

The Nuclear Photoelectric Effect

25. We have discussed the photoelectric effect, in which radiation falling on an atom ejected electrons from the structure of the atom. We found that with radiation of frequency ν is associated energy given by $E = h\nu$, where h is Planck's constant. It might be supposed that if radiation of energy greater than the binding energy of a nucleus fell on that nucleus, disintegration would occur, the particles composing the nucleus being released from the grip of the binding force and flying apart. This proves experimentally to be true. Investigations of this sort are limited by the fact that the most energetic gamma radiation yet found in nature has an energy of 2.6 mev, whereas the binding energies of all but two kinds of nuclei are greater than this. The two nuclei that can be disintegrated by gamma radiation are the deuteron, which has a binding energy around 2 mev, and ${}_4Be^9$, whose binding energy is about 1.3 mev. The production in nuclear reactions of gamma radiation of energy as high as 16 mev takes place, and many other nuclei have been disintegrated by this radiation.

26. Nuclear transmutations such as the photodisintegration of deuterium are represented by equations similar to those used to represent chemical reactions. For example,

$$ {}_1H^2 + h\nu = {}_1H^1 + {}_0n^1 + \text{K.E.} \tag{6} $$

symbolizes the reaction between the deuteron and a photon of energy $h\nu$. K.E. stands for the sum of the kinetic energies of the proton and neutron produced by the disintegration, whose mass equivalent must be taken into account in balancing the reaction. From a knowledge of two of the three atomic masses entering into the above reaction and a knowledge of the energy of the radiation and the kinetic energy of the products, the mass of the third particle can be determined quite exactly. It is for the purpose of calculating as exactly as possible the masses of light atoms (and hence learning their binding energies) that such equations have been chiefly applied. Suppose that we wished to find the mass of the neutron by studying the photodisintegration of the deuteron and we had the following data: mass of ${}_1H^1 = 1.0081$, mass of ${}_1H^2 = 2.0147$, energy of gamma radiation $h\nu = 2.6$ mev, sum of K.E. of neutron and proton = 0.5 mev.

The atomic masses we have obtained from mass-spectrographic data, the energy of the gamma radiation from measurements on the radioactive body used as a source in the experiment, and the K.E. of the products have been measured in the photodissociation experiment. Solving the foregoing equation for the mass of the neutron, we have

$$_0n^1 = {}_1H^2 - {}_1H^1 + \frac{h\nu - \text{K.E.}}{k} \tag{7}$$

where k is the constant necessary to transform mass units into mev. Then,

$$\begin{aligned} _0n^1 &= (2.0147 - 1.0081) + \frac{2.6 - 0.5}{930.95} \\ &= 1.0066 + \frac{2.1}{930.95} \\ &= 1.0089. \end{aligned}$$

Nuclear Disintegration by Particle Bombardment

27. It is possible to produce nuclear disintegration and transmutation by means of bombardment with swiftly moving nuclei as well as by irradiation with high-energy gamma rays. The simplest nuclei are usually employed as projectiles: protons, deuterons, and alpha particles. The last named are convenient projectiles because they are emitted with high energies from radioactive substances, and nuclear transmutation by alpha-particle bombardment was historically the first to be discovered. In 1919, Rutherford found that certain elements, when bombarded by alpha particles, emitted swift protons, the alpha particle actually being absorbed into the nuclear structure and a proton being emitted. Photographs of such events were obtained by the cloud-track method, so that there remained no doubt about what had happened. A typical example of such a disintegration is that of nitrogen by alpha particles:

$$_7N^{14} + {}_2He^4 = {}_8O^{17} + {}_1H^1 + \text{K.E.}_{\text{H}} - \text{K.E.}_{\text{He}}. \tag{8}$$

It is evident that the kinetic energy possessed by the alpha particle which produces the disintegration must be taken into account in balancing the equation for the reaction.

28. The sort of nuclear transmutation discussed in the last section requires particles emitted by a naturally radioactive sub-

stance. In 1932, Cockcroft and Walton, working in the Cavendish Laboratory, succeeded in producing the first wholly "man-made" disintegration. They accelerated protons in a discharge tube by means of a potential of about 600,000 volts and allowed these swift protons to bombard a target of lithium (Fig. 291). Alpha particles were emitted from the target under the bombardment. The pioneer work of Cockcroft and Walton was

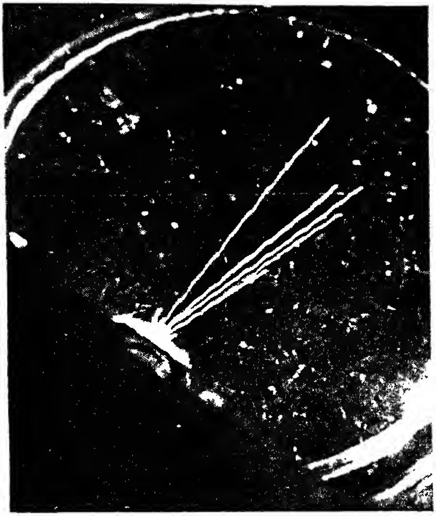


FIG. 291.—Alpha particles from lithium that is being bombarded with protons. ($_3Li^7 + _1H^1 \rightarrow 2\,_2He^4$). (*Delsasso, Fowler, and Lauritsen.*)

immediately followed by the production of nuclear transmutations in several laboratories in this country and in Europe, and the development of nuclear physics has progressed with extreme rapidity. The deuteron was found to be a much more effective projectile than the proton for producing disintegrations, and voltages available for the acceleration of bombarding particles have continually increased.

The Cyclotron

29. The study of nuclear transmutation demands some means of getting high-energy particles. It is interesting that the two most successful means for getting such particles have both been

developed in this country and are so simple in principle that they can be understood by beginning students. Both methods depend on the acceleration of charged atoms by electric fields. In one, the ions are accelerated by a continuous fall through a very high potential drop provided by a Van de Graaff generator (see Chap. X, Par. 20). Such a potential may alternatively be obtained by

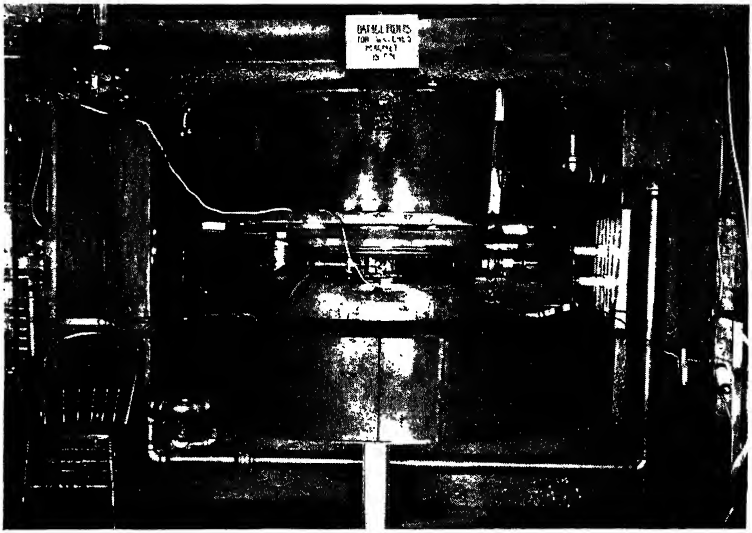


FIG. 292.—A cyclotron with the box containing the "D's" in place between the pole pieces of the magnet. The pumps, oscillator and leads to the "D's" are in back out of sight. (*Palmer Laboratory, Princeton.*)

transformers and rectifiers but, for potentials over a million volts, a Van de Graaff machine has proved much cheaper and simpler. In the other method, the same ions fall repeatedly through a comparatively small voltage, increasing their kinetic energy with each passage until it has reached an amount corresponding to a single voltage drop of millions of volts. The device which makes this possible is called a cyclotron and was invented by E. O. Lawrence at the University of California.

30. The main apparent feature of a cyclotron, as may be seen in Fig. 292, is an enormous electromagnet, and yet in a way this is a piece of auxiliary equipment. Its function is merely to keep

bending the speeding ions around in semicircles so that they can be further accelerated. The principle involved is the one so familiar in the study of cathode rays and positive rays, namely, that a charged particle moving perpendicularly to a magnet field travels in a circular path. (This principle is nicely illustrated by one of the tracks in Fig. 293.) To understand the details of the

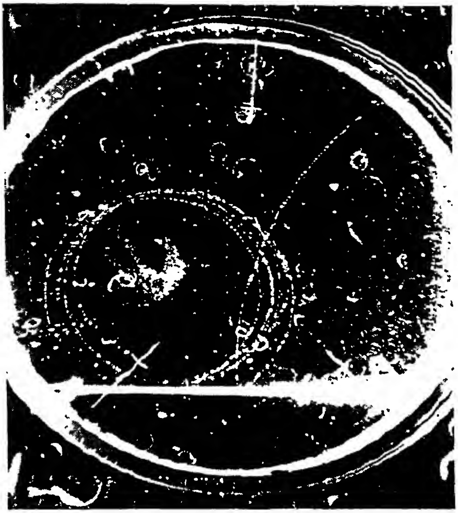


FIG. 293.—Tracks of an electron-positron pair in which one of the particles has made several complete circles under the influence of a magnetic field perpendicular to the plane of the photograph. (*Delsasso, Fowler, and Lauritsen.*)

process, consider Figs. 294, 295, and 296. Of these, the first two are photographs of the core of the apparatus in two stages of assembly, and Fig. 296 is a simplified schematic drawing of the region between the pole pieces of the magnet. D_1 and D_2 are flat semicircular copper boxes open along their straight sides; they are like the two halves of a pancake-shaped metal box that has been cut along a diameter. These "D's" are mounted on insulating supports in a circular evacuated box that fits between the poles of the magnet. Each D is connected to one terminal of a high-frequency generator of alternating voltage (about 50,000 volts). Since the D's are of metal, there is never an electric field inside them, but in the region between them there is an alternating

electric field directed first toward one and then toward the other. When the D's are between the poles of the magnet, there is a

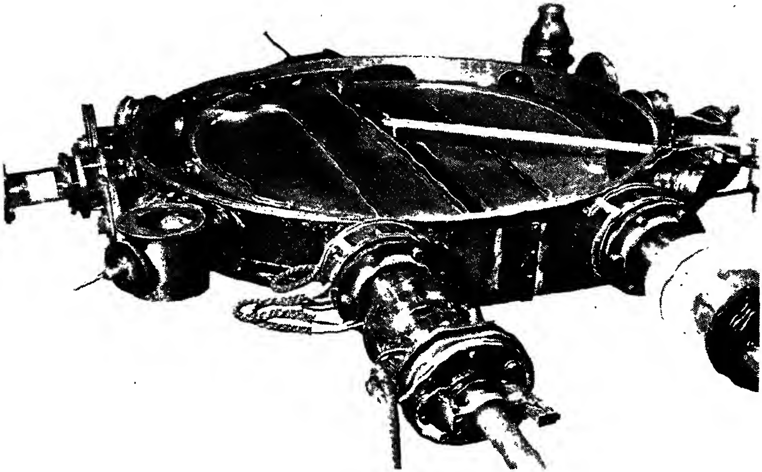


FIG. 294.—Cyclotron box with lid removed, showing D's. The deflecting plate may also be seen close to outside of the left D.

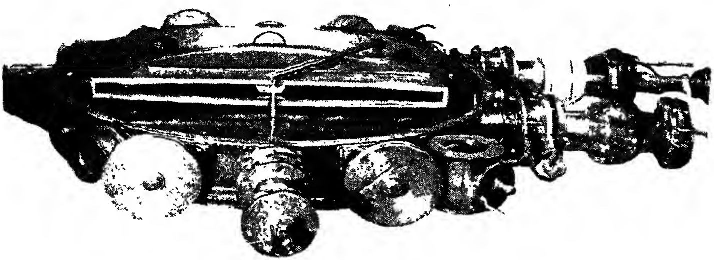


FIG. 295.—The same as Fig. 294 but with the left D removed, showing the structure of the D's and the ion source in the center.

magnetic field everywhere in them and between them, in a direction perpendicular to their plane. A filament is mounted over A, the center of the space between the D's, and electrons from it

are shot down along the direction of the magnetic field. Although the gas pressure is very low in the chamber, there is a constant stream of gas flowing in and being pumped out so that there are always some molecules in the path of the electrons, and some of these molecules are ionized by collision with the electrons. Thus, positive ions are formed in the region A, but everywhere between the D's there is an alternating electric field. This field accelerates the ions toward one of the D's, the one that happens at the moment to be negative. Once inside the D, the ions are

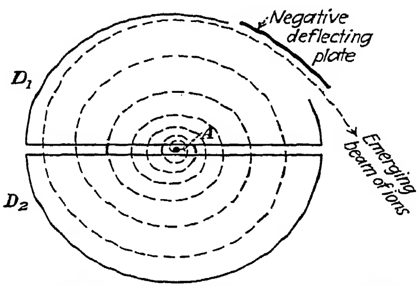


FIG. 296.—Schematic diagram of cyclotron D's, indicating the spiral path of the ions. The true path has many more turns in the spiral than are shown.

shielded from the electric field but are still in a magnetic field perpendicular to the velocity they have just acquired. Consequently, they will move in a semicircular path and return to the region between the D's where they are once more subject to the electric field. They have lost no energy in their semicircular trip through the D and are still moving with the velocity resulting from their first acceleration by the electric field, but time has elapsed and it may be that the direction of the electric field has changed. If so, the ions are again accelerated, gaining more kinetic energy before passing into the second D. Here they again travel in a semicircle, but one of larger radius, since they are going faster. They again come to the space between the D's, speed up some more, and go back into the first D. This process continues with the ions traveling faster and faster in greater and greater semicircles until they reach the outer edge of one of the D's where an auxiliary electrode at a negative potential pulls them out into a tangential path which leads them through a thin win-

dow of metal foil into a chamber on the outside of the box. Here they strike the target that is being investigated.

31. Evidently, the crucial point of this apparatus is that the alternating electric field should change direction in such a way that it always accelerates the ions, never slows them. Fortunately this condition is easily satisfied as we can see by calculating the time an ion spends in the D on each semicircular segment of its path. From Chap. XX, Par. 5, we know that the radius of curvature of an ion in a magnetic field perpendicular to its direction of motion is

$$R = \frac{mv}{Be}. \tag{9}$$

On each semicircular trip through a D, the ion has to travel the distance πR, and therefore the time T required for such a trip is $T = \pi R/v$ or, by substituting for R,

$$T = \frac{\pi}{v}\frac{mv}{Be} \qquad \frac{\pi m}{Be}. \tag{10}$$

This last expression for T does not contain either the velocity or the radius; evidently the faster the ions go, the farther they go, and the two effects cancel each other. The fast ions, then, take just as long to get around through the D's as the slow ions do. This means that if the period of oscillation of the electric field is $2T = 2\pi m/Be$ the ions will be speeded up every time they come around. The limit to the energy that can be obtained is the limit to the size of the pole pieces of the magnet and to the magnetic induction that can be obtained in the region between them. With the values that are practically attainable, T is so small that the alternating electric field has to be supplied by a short-wave oscillator with a frequency range of 10 to 20 megacycles/sec., corresponding to a wave length of from 15 to 30 m.

32. The highest energies ever imparted to atomic particles by artificial means have been obtained with the cyclotron. For protons, the maximum reported up to the summer of 1939 is 7.5 million volts (for deuterons 19.0 million volts). The currents of ions produced at these high voltages are naturally very small, of the order of a few microamperes.

33. The necessity of employing high-energy projectiles to produce nuclear disintegration is, of course, a result of the fact

that the bombarding nuclei, like the target nuclei, are positively charged. Before the bombarding nuclei can penetrate to the region where the intranuclear binding force operates—to which region they must penetrate before a disintegration can take place—they must overcome the electrostatic repulsion that they suffer at distances greater than about 3×10^{-15} m. from the nucleus. It is also evident that with increasing atomic number (*i.e.*, with increasing charge on the nucleus) it will become necessary that the particles have higher and higher energies in order to produce disintegration. With the 19 mev deuterons already mentioned, transmutations in most of the elements have been produced. With energies of 1 mev and less, one is more or less restricted to the study of reactions in elements of atomic number 11 or smaller. When neutrons are employed as the projectiles in nuclear bombardment, however, the foregoing considerations do not apply, since the neutron has no charge and there is consequently no electrostatic repulsion between it and the target nucleus. It is experimentally found, as a matter of fact, that very slow neutrons are much more effective in the production of a certain reaction (4*c*, below) than are swift neutrons; and the charge number of the bombarded nucleus seems to have no such influence on the probability of disintegration by neutron bombardment as it has in the case of bombardment by charged particles.

Classification of Nuclear Reactions

34. Almost without exception, any of the known nuclear reactions is of one or another of the following types:

1. When nuclei are bombarded with alpha particles, the result may be
   a. Capture of the alpha particle and emission of a swift proton.
   b. Capture of the alpha particle and emission of a neutron.
2. When nuclei are bombarded with protons, the result may be
   a. Capture of the proton and emission of an alpha particle.
   b. Capture of the proton accompanied by gamma-ray emission.
   c. Capture of the proton and emission of a neutron.

3. When nuclei are bombarded with deuterons, the result may be
   a. Capture of the deuteron and emission of an alpha particle.
   b. Capture of the deuteron and emission of a proton.
   c. Capture of the deuteron and emission of a neutron.
4. When nuclei are bombarded with neutrons, the result may be
   a. Capture of the neutron and emission of an alpha particle.
   b. Capture of the neutron and emission of a proton.
   c. Capture of the neutron accompanied by gamma-ray emission.
   d. Capture of the neutron and emission of two neutrons.

One example will enable the student to work out for himself examples of the other type reactions. It is easy to see that reaction 3*c* will result in a nucleus of mass number one greater and charge number one greater than the target nucleus, that, in effect, a proton has been added to the target nucleus. When deuterons bombard carbon, reaction 3*c* takes place:

$$ {}_6C^{12} + {}_1H^2 = {}_7N^{13} + {}_0n^1 + K.E._n - K.E._H. \qquad (11) $$

Artificial Radioactivity

35. Some students may recognize that nitrogen has no known stable isotope of mass number 13. In fact, the nitrogen produced in the reaction of Eq. (11) is unstable. It disintegrates radioactively with a half-life of 10.3 min., emitting positrons and becoming carbon 13, a stable isotope of ordinary carbon. We can then explain the fact that N^{13} is not found in nature on the ground of its instability. Many such unstable isotopes of the familiar elements are formed in the course of nuclear reactions, some of them emitting negative electrons to return to a stable form, and some of them emitting positrons. The study of these "artificial" radioactive elements has led and is leading to valuable advances in our understanding of nuclear phenomena.

SUMMARY

The electromagnetic-wave theory of light was triumphantly successful in explaining the phenomena known to nineteenth-

century physics, but in more recent years it has encountered difficulties. Electromagnetic waves arise whenever electric charges are accelerated. The electrons moving in orbits around the nuclei of atoms are accelerated. Therefore they should gradually draw closer and closer to the nucleus, constantly sending out electromagnetic radiation in the process. To explain why this does not happen, we assume that there are only certain electron orbits permissible and that not more than one electron is ever in a given orbit. Radiation occurs when an electron jumps from one orbit to another closer to the nucleus. The number and position of permissible orbits depend on the charge on the nucleus, *i.e.*, the atomic number. This dependence explains the differences in the spectra given off and absorbed by different elements.

The photoelectric effect is even more at variance with nineteenth-century wave theory than is atomic structure. The energy of an electron released from a metal surface by light depends not on the intensity of the light but on its frequency. The kinetic energy of photoelectrons is given by Einstein's equation

$$\tfrac{1}{2}mv^2 = h\nu - \phi.$$

Apparently, light behaves in some ways as if it were composed of corpuscles each containing energy $h\nu$ and momentum $h\nu/c$. These are called photons. The existence of photons is also required to explain the Compton effect in x-rays.

Not only do light waves have corpuscular properties, but particles have wave properties. Electrons are diffracted as if they had a wave length h/mv. Apparently the concepts of wave and particle lose their distinction when applied to submicroscopic particles. There is one set of laws, the laws of wave mechanics, that explains both types of phenomenon.

The neutron is a particle of about the mass of a proton but carrying no charge. It has no direct ionizing action and therefore cannot be detected directly in a Wilson cloud chamber, but it may collide with nuclei which then make visible tracks. Atomic nuclei appear to be made up of protons and neutrons held together by attractive forces that predominate over Coulomb forces at distances of less than 10^{-14} m. Protons, deuterons, and alpha particles can penetrate atomic nuclei if they have sufficient

kinetic energy. Such penetration results in the formation of different nuclei. These new nuclei formed by transmutation are sometimes stable known atomic species and sometimes unstable previously unknown atomic species that disintegrate radioactively. Nuclear transformations can also be effected by high-energy gamma rays and by neutrons. Many of these artificial radioactive elements give off positrons when they revert to a stable form. Positrons are of approximately the same mass and charge as electrons, but their charge is positive.

Two means of producing very high energy particles for causing nuclear disintegrations have been used successfully. In one, the Van de Graaff generator produces a very high potential difference which accelerates positive ions in a discharge tube. In the other, the cyclotron, ions are given high energies by repeated accelerations in comparatively small fields. In the cyclotron, a magnetic field bends the ions in semicircular paths, bringing them back repeatedly to a region between two electrodes where they are accelerated.

ILLUSTRATIVE PROBLEMS

1. The mercury line of wave length $2{,}536 \times 10^{-10}$ m. is used to eject photoelectrons from silver. The difference of potential required to bring the electrons to rest is 0.11 volt. Find the energy necessary to remove an electron from silver.

The energy available in a photon of $2{,}536 \times 10^{-10}$ m. wave length is, Par. 11,

$$h\nu = h\frac{c}{\lambda} = \frac{6.55 \times 10^{-34} \text{ joule-sec. } 3 \times 10^{8} \text{ m./sec.}}{2{,}536 \times 10^{-10} \text{ m.}}$$
$$= 7.75 \times 10^{-19} \text{ joule.}$$

The energy necessary to remove an electron from silver is, Eq. (1), page 590,

$$\varphi = h\nu - \tfrac{1}{2}mv^2. \tag{12}$$

Since 0.11 volt is required to stop the electrons, their initial kinetic energy as they left the silver surface was Eq. (9), page 395,

$$\tfrac{1}{2}mv^2 = Ve = 0.11 \text{ volt} \times 1.60 \times 10^{-19} \text{ coulomb}$$
$$= 1.76 \times 10^{-20} \text{ joule.}$$

Therefore, substituting the value of $h\nu$ found previously and this value of the kinetic energy in Eq. (12), we find

$$\phi = (7.75 \times 10^{-19} - 1.76 \times 10^{-20}) \text{ joule} = 7.57 \times 10^{-19} \text{ joule}$$

$$\phi = \frac{7.57 \times 10^{-19}\text{ joule}}{1.60 \times 10^{-19}\text{ coulomb}} = 4.73 \text{ electron volts.}$$

2. X-rays of 100 X units wave length fall on a carbon scatterer. The scattered radiation consists of x-rays of 112 X units wave length. Find the directions of the scattered photon and the recoil electron.

The angle that the scattered photon makes with the path of the incident photon is, by Eq. (5), page 594,

$$\lambda' - \lambda = \frac{h}{m_0 c}(1 - \cos\varphi)$$

or

$$\cos\varphi = 1 - \frac{m_0 c}{h}(\lambda' - \lambda) \tag{13}$$

where $m_0 = 9.04 \times 10^{-31}$ kg. = the mass of a stationary electron.
$h = 6.55 \times 10^{-34}$ joule-sec. = the Planck constant.
$c = 3 \times 10^8$ m./sec. = the velocity of light.
$\lambda' = 112$ X units $= 112 \times 10^{-13}$ m. = the wave length of the scattered photon.
$\lambda = 100$ X units $= 100 \times 10^{-13}$ m. = the wave length of the incident photon.

Substituting these values in Eq. (13), we find

$$\cos\varphi = 1 - \frac{9.04 \times 10^{-31}\text{ kg.} \times 3 \times 10^8 \text{ m./sec.}}{6.55 \times 10^{-34}\text{ joule-sec.}}(112 - 100) \times 10^{-13}\text{ m.}$$

$$\cos\varphi = 1 - \frac{9.04 \times 3 \times 12}{6.55}10^{-2} = 1 - 0.497 = 0.503$$

$$\varphi = \arccos 0.503 = 59°48'$$

$$\sin\varphi = 0.864.$$

The angle θ which the recoil electron makes with the path of the incident photon may be found from Eqs. (3) and (4), see pages 593 and 594.

$$\frac{h\nu}{c} = mv\cos\theta + \frac{h\nu'}{c}\cos\varphi$$

$$0 = mv\sin\theta - \frac{h\nu'}{c}\sin\varphi.$$

Solving each equation for the term containing θ, we obtain

$$mv\sin\theta = \frac{h\nu'}{c}\sin\varphi$$

$$mv\cos\theta = \frac{h\nu}{c} - \frac{h\nu'}{c}\cos\varphi.$$

Then if we divide the first of these equations by the second, we get

$$\tan\theta = \frac{\dfrac{h\nu'}{c}\sin\varphi}{\dfrac{h\nu}{c} - \dfrac{h\nu'}{c}\cos\varphi} = \frac{\dfrac{1}{\lambda'}\sin\varphi}{\dfrac{1}{\lambda} - \dfrac{1}{\lambda'}\cos\varphi} = \frac{\lambda\sin\varphi}{\lambda' - \lambda\cos\varphi}$$

since $\nu = c/\lambda$ and $\nu' = c/\lambda'$. Therefore,

$$\tan\theta = \frac{100 \times 10^{-13}\text{ m. } 0.864}{112 \times 10^{-13}\text{ m.} - 100 \times 10^{-13}\text{ m. } 0.503} = \frac{86.4}{61.7}$$

$$\theta = \arctan 1.400 = 54°28'.$$

3. Alpha particles are emitted when nitrogen is bombarded with deuterons. The energy of the incident deuterons was 0.575 mev. Find the energy of the alpha particles emitted. Assume that all the excess energy goes into the alpha particles.

The equation of this reaction is

$${}_7N^{14} + {}_1H^2 = {}_6C^{12} + {}_2He^4 + E_2 - E_1.$$

The energy of the incident deuterons is

$$E_1 = 0.575\text{ mev} = \frac{0.575\text{ mev}}{931\text{ mev/m. u.}} = 6.18 \times 10^{-4}\text{ m. u.}$$

If we write in the masses from Table 17, Chap. XX, we obtain

$$14.0073\text{ m.u.} + 2.0147\text{ m.u.} = 12.0036\text{ m.u.} + 4.0039 + E_2 - 0.0006\text{ m.u.}$$

From this equation,

$$\begin{aligned} E_2 = 0.0151\text{ m.u.} &= 0.0151\text{ m.u.} \times 931\text{ mev/m.u.} \\ &= 14.1\text{ mev} \\ &= \text{the energy of the } \alpha \text{ particles emitted.} \end{aligned}$$

PROBLEMS

1. The energies of the different stationary states of the hydrogen atom were given by Bohr as

$$E = -\frac{2\pi^2 m e^4}{(4\pi\epsilon_0)^2 h^2 n^2}$$

where m is the mass and e the charge of the electron, h Planck's constant, ϵ_0 the dielectric constant for a vacuum, and n an integer 1, 2, 3, . . . for successive states. Find the wave lengths of the first three lines of the Balmer series which starts with $n = 2$ as the lower level, and compare them with the observed values.

2. The energy necessary to remove an electron from metallic silver is 4.73 electron volts. Compute the threshold wave length, *i.e.*, the wave length just short enough to remove electrons from silver.

3. Under a certain illumination, the electrons emitted by cobalt for which the threshold wave length is $6{,}600 \times 10^{-10}$ m., require 100 volts opposing potential to bring them to rest. What is the frequency of the light used to eject these electrons?

4. The velocity of the electrons emitted by a sodium surface is 4×10^6 m./sec. If the frequency of the light used to eject these electrons is 7.1×10^{14} vibrations per second, what is the threshold wave length?

5. Show that the change in wave length in the Compton effect is

$$\lambda' - \lambda = \Delta\lambda = \frac{h}{mc}(1 - \cos \varphi).$$

Use the kinetic energy and momentum of the emitted electron in the relativistic form given in the footnote to page 594.

6. X rays of wave length 100 X units fall on a carbon scatterer. The scattered radiation consists of x-rays of wave length 124 X units. Find the speed of the recoil electron.

7. What is the change in the wave length of x-rays scattered through 90°?

8. X-rays of 100 X units are scattered through an angle of 90°. What is the angle between the path of the recoil electron and that of the original photon.

9. If the wave length of the regulation league baseball traveling 30 mi./hr. is about 3.5×10^{-35} m., what is the mass of the ball?

10. What acceleration would be produced in a proton by a force of 230 newtons?

11. Compute the binding energy of the ${}_3Li^7$ nucleus.

12. Show that the energy required to create an electron is 0.5 mev.

13. How large a diameter magnet, producing a field of 1.6×10^6 amp. turns/m., would a cyclotron have to have to yield 10 mev protons?

14. In the disintegration of nitrogen, alpha particles with a velocity of 6.05×10^5 m./sec. from Ra C′ were used. Compute the kinetic energy of the proton emitted in the reaction.

15. Alpha particles are emitted when lithium is bombarded with protons. It is known that they come from ${}_3Li^7$. Write down the equation for this nuclear reaction, and calculate the energy of each alpha particle. Assume the energy divided equally between the two alpha particles produced, and assume an initial proton energy of 0.19 mev.

16. In the disintegration of ${}_6C^{12}$ by deuterons, the kinetic energy of the deuterons is 0.28 mev, and the kinetic energy of the neutron emitted is negligible. Find the mass of N^{13}.

APPENDIX

I. NUMERICAL DATA

TABLE 1.—ELASTIC MODULI
Average values of elastic constants in newtons per sq. m.
1 newton/sq. m. = 10 dynes/sq. cm.

Material M	Bulk modulus k	Modulus of rigidity n	Young's modulus y
Aluminum	7.6×10^{10}	2.4×10^{10}	6.0×10^{10}
Copper	14.3×10^{10}	4.2×10^{10}	12.0×10^{10}
Steel	18.1×10^{10}	8.0×10^{10}	20.0×10^{10}
Lead	5.0×10^{10}	0.54×10^{10}	1.6×10^{10}
India rubber	....	0.00016×10^{10}	0.05×10^{10}
Crown glass	5.0×10^{10}	2.9×10^{10}	7.1×10^{10}
Oak	....	....	1.3×10^{10}

TABLE 2.—SURFACE TENSIONS OF LIQUIDS IN CONTACT WITH AIR AT SPECIFIED TEMPERATURES
1 newton/m. = 10^3 dynes/cm.

Substance	Newtons/m.	°C.
Aluminum (molten)	0.520	750
Benzene	0.0289	20
Carbon tetrachloride	0.0268	20
Ether	0.0170	20
Ethyl alcohol	0.02227	20
Hydrogen	0.00191	−252.
Lead (molten)	0.422	350
Mercury	0.476	20
Soap solution	0.028	20
Water	0.07275	20

TABLE 3.—COEFFICIENTS OF DYNAMIC FRICTION

Glass on glass	0.40	Brass on brass	0.20
Mild steel on mild steel	0.56	Steel on ice	0.02
Hard steel on hard steel	0.42	Rubber on dry concrete	0.5–0.8
Nickel on mild steel	0.65	Leather on metals	0.5–0.6
Carbon on glass	0.18	Leather on metals, oiled	0.15
Wood on wood	0.25–0.50	Lubricated metal surfaces	0.05–0.2

TABLE 4.—COEFFICIENTS OF VISCOSITY

Substance	Coefficient of viscosity, m.k.s. units*	Temperature, °C.
Air	0.0182×10^{-3}	20
Ethyl alcohol	1.1943×10^{-3}	20
Ethyl ether	0.234×10^{-3}	20
Glycerin	$830. \times 10^{-3}$	20
Hydrogen	0.0097×10^{-3}	20
Mercury	1.56×10^{-3}	20
Oil, castor	$986. \times 10^{-3}$	20
Oil, machine	$100–600 \times 10^{-3}$	20
Oil, olive	$84. \times 10^{-3}$	20
Water	1.7938×10^{-3}	0
	1.0087×10^{-3}	20
	0.2839×10^{-3}	100

* The m.k.s. unit of viscosity is 1 newton-sec./sq. m. This unit is ten times as large : the c.g.s. unit, the dyne-second per square centimeter.

TABLE 5.—COEFFICIENTS OF CUBICAL THERMAL EXPANSION

$$V_t = V_0(1 + \alpha t + \beta t^2 + \gamma t^3).$$

Substance	Temperature range, °C.	α	β	γ
Ethyl alcohol	0– 80	1.0414×10^{-3}	0.7836×10^{-6}	1.7168×10^{-8}
Mercury	24–299	0.18182×10^{-3}	0.00078×10^{-6}	
Pyrex glass	21–471	0.0108×10^{-3}		
Water	0– 33	-0.0643×10^{-3}	8.505×10^{-6}	6.790×10^{-8}

TABLE 6.—COEFFICIENTS OF LINEAR THERMAL EXPANSION

Change in length per unit length per degree Centigrade, *i.e.*, $L_t = L_0(1 + at)$.

	a in $\frac{\text{m.}}{\text{m. °C}}$
Aluminum	2.394×10^{-5}
Brass	1.818×10^{-5}
Copper	1.771×10^{-5}
Glass, flint	0.78×10^{-5}
Hard	0.97×10^{-5}
Pyrex	0.3×10^{-5}
Soft	0.85×10^{-5}
Gutta-percha	19.8×10^{-5}
Platinum	0.89×10^{-5}
Platinum-iridium	0.87×10^{-5}
Quartz	0.058×10^{-5}
Steel	1.332×10^{-5}
Tungsten	0.444×10^{-5}
Wood along the grain, mahogany	0.3×10^{-5}
Pine	0.5×10^{-5}
Wood across the grain, mahogany	4.0×10^{-5}
Pine	3.4×10^{-5}

TABLE 7.—SPECIFIC HEATS

Substance	Specific heat	Temperature, °C.
Aluminum	0.217	17–100
Copper	0.0931	15–100
Glass crown	0.161	10– 50
Flint	0.117	10– 50
Thermometer	0.199	19–100
Iron	0.114	20–100
Lead	0.031	20–100
Lithium	0.941	
Mercury	0.0331	40
Silver	0.0570	
Sulfur	0.176	
Tin	0.0562	
Water, ice	0.502	0
Liquid	1.0049	4
	1.0000	15
	0.9971	30
Wood	0.42	
Zinc	0.0935	17–100

TABLE 8.—MOLAR HEATS OF GASES IN *R* UNITS

	c_V	c_P	$c_P - c_V$	c_P/c_V
Monatomic gases:				
Theoretical	1.5	2.5	1.0	1.67
Helium	1.51	2.50	0.99	1.66
Neon	1.50	2.50	1.01	1.66
Argon	1.50	2.46	0.94	1.64
Diatomic gases:				
Theoretical	2.5	3.5	1.0	1.40
Air	2.56	3.60	1.04	1.41
Hydrogen	2.44	3.47	1.03	1.42
Polyatomic gases:				
Theoretical				
Carbon dioxide	3.63	4.84	1.21	1.33
Water vapor	3.08	4.21	1.13	1.37

TABLE 9.—HEATS OF VAPORIZATION

Substance	Kg.-cal./kg.	
Carbon dioxide	87.2	−60
	71.4	−30
Carbon tetrachloride	46.4	76.8
Copper	,755	2,300
Ethyl alcohol	204	78.4
Ethyl ether	83.9	34.6
Helium	6	−268.9
Hydrogen	108	−252.7
Iron	,626	3,000
Lead	222.6	1,620
Mercury	70.8	357
Water	595.9	0
	568.5	50
	539.6	100
	503.5	150
Zinc	362.5	907

TABLE 10.—HEATS OF FUSION

Substance	Kg.-cal./kg.	At °C.
Carbon dioxide	45.30	− 56.2
Copper.......	43.2	1083
Ethyl alcohol..	24.9	− 112
Iron..........	48.0	1535
Lead.........	5.86	327
Tin..........	14	231.8
Water........	79.67	0
Zinc..........	26.1	419.4

TABLE 11.—CRITICAL DATA

Substance	Critical temperature, °C.	Critical pressure, atm.	Density, g./cc.
Air....................	-140.7	37.2	0.31
Argon..................	-122	48	0.531
Benzene................	288.5	47.7	0.304
Carbon dioxide..........	31.1	73.0	0.460
Carbon disulfide.........	273	76	
Carbon monoxide........	-139	35	0.311
Chlorine................	144.0	76.1	0.573
Ethyl alcohol...........	243.1	63.1	0.2755
Helium.................	-267.9	2.26	0.066
Hydrogen..............	-239.9	12.8	0.0310
Methane...............	- 82.5	45.8	0.162
Neon..................	-228.7	25.9	0.484
Nitric oxide............	- 94	65	0.52
Nitrogen...............	-147.1	33.5	0.311
Oxygen................	-118.8	49.7	0.430
Sulfur dioxide...........	157.2	77.7	0.52
Water..................	374.0	217.72	0.4

TABLE 12.—RELATIVE DIELECTRIC CONSTANTS

Multiply by $\epsilon_0 = 8.84 \times 10^{-12}$ farad/m. to get dielectric constants in m.k.s. practical units.

Substance	K	Substance	K
Air	1.00059	Mica	5.6–6.0
Asphalt	2.68	Oil, insulating	2.5
Benzene	2.3	Paper	3.5
Bromine	3.18	Paraffin	2.10
Diamond	16.5	Porcelain	6
Ebonite	2.72	Shellac	3.1
Ethyl alcohol	25.8	Sulfur	4.2
Ethyl ether	4.3	Water	81.1
Glass (flint)	9.90	Wood	3
Ice (−2°C.)	2.9		

TABLE 13.—THERMOELECTRIC E.M.FS.

Millivolts for several pairs of metals with the hot junction at a temperature t°C. and the cold junction at 0°C.

t	Ag-Cu	Ni-Cu	Fe-Cu	Ag-Pt	Chromel-alumel
100	−0.04	− 2.26	+1.06	+0.72	4.08
200	−0.009	− 4.89	+1.52	+1.73	8.17
300	−0.17	− 7.56	+1.38	+2.96	12.32
400	−0.17	− 9.74	+0.64	+4.47	16.50
500	−0.13	−11.94	−0.70	+6.26	20.76

The plus sign indicates current flowing from the hot to the cold junction in the first metal of the pair.

TABLE 14.—RESISTIVITY

$$\rho = \frac{AR}{l}$$

Substance	Ohm-m. at 20°C.
Aluminum	2.62×10^{-8}
Copper	1.77×10^{-8}
Gold	2.44×10^{-8}
Iron	10×10^{-8}
Lead	22×10^{-8}
Manganin	44×10^{-8}
Nichrome	100×10^{-8}
Platinum	10×10^{-8}
Silver	1.62×10^{-8}

TABLE 15.—TEMPERATURE COEFFICIENTS OF RESISTANCE

Change of resistance per unit resistance per degree change in temperature.

$R = R_0(1 + \theta T)$.

Substance	θ	Temperature range, °C.
Aluminum	3.8×10^{-3}	18–100
Copper	4.3×10^{-3}	18
Gold	4×10^{-3}	0–100
Lead	4.3×10^{-3}	18
Manganin	$2 \times 10^{-6} - 5 \times 10^{-5}$	20
Nichrome	1.7×10^{-4}	20
Platinum	3.8×10^{-3}	0–100
Silver	4.0×10^{-3}	0–100

TABLE 16.—MOMENTS OF INERTIA

Body	Axis	Moment of inertia
Uniform thin rod.	Normal to length at one end	$\frac{ML^2}{3}$
Uniform thin rod.	Normal to length at center	$\frac{ML^2}{12}$
Thin circular sheet of radius R	Normal to the plate through the center	$\frac{MR^2}{2}$
Thin circular sheet of radius R	Along any diameter	$\frac{MR^2}{4}$
Rectangular parallelopiped, edges a, b, c.	Through center perpendicular to face ab	$\frac{M(a^2 + b^2)}{12}$
Sphere of radius R.	Any diameter	$\frac{2MR^2}{5}$
Right circular cylinder of radius R and length L	Longitudinal axis of cylinder	$\frac{MR^2}{2}$
Right circular cylinder of radius R and length L	Through center perpendicular to axis of cylinder	$M\left(\frac{R^2}{4} + \frac{L^2}{12}\right)$

TABLE 17.—MASS NUMBERS AND PACKING FRACTIONS OF ATOMIC SPECIES

(See text, pages 419–421.)

TABLE 18.—RADIOACTIVE DATA

Element uranium series	Symbol	At. no.	At. wt.	Half life	Radiation	Range in air, cm.	Velocity in fractions of speed of light
Uranium I.......	UI	92	238	4.5×10^9 yr.	α	2.37	0.0456*
Uranium X_1.....	U-X_1	90	234	23.8 days	β		
Uranium X_2.....	U-X_2	91	234	1.15 min.	$\beta(\gamma)$		
Uranium II......	UII	92	234	2×10^6 yr.	α	2.75	0.0479*
Ionium..........	Io	90	230	6.9×10^4 yr.	α	2.85	0.0485*
Radium.........	Ra	88	226	1,580 yr.	$\alpha(\beta, \gamma)$	3.13	0.0500*
Radon..........	Rn	86	222	3.85 days	α	3.94	0.054*
Radium A.......	Ra-A	84	218	3.0 min.	α	4.50	0.0565*
Radium B.......	Ra-B	82	214	26.8 min.	$\beta(\gamma)$	....	0.74* (max.)
Radium C.......	Ra-C	83	214	19.5 min.	99.97 % β, γ	....	0.96* (max.)
Radium C'....	Ra-C'	84	214	10^{-6} sec.	α	6.57	0.0641*
Radium D (Radiolead)..	Ra-D	82	210	16.5 yr.	β, γ	....	0.39* (max.)
Radium E.....	Ra-E	83	210	5.0 days	β		
Radium F (Polonium)..	Ra-F	84	210	136 days	$\alpha(\gamma)$	3.58	0.0523*
Radium G.....	Ra-G	82	206				
Radium C.......	Ra-C	83	214	....	0.03 % α		
Radium C''....	Ra-C''	81	210	1.4 min.	β		

* To get velocity in meters per second multiply by 3×10^8.

For similar data on the actinium series and the thorium series see the International Table of the Radioactive Elements and Their Constants in the "Handbook of Chemistry and Physics" of the Chemical Rubber Company.

TABLE 19.—INDICES OF REFRACTION

Air	1.000294
Benzene	1.495
Carbon tetrachloride	1.46
Glass, silicate flint	1.5794
Borosilicate flint	1.5503
Heavy flint	1.9625
Silicate crown	1.4782
Borosilicate crown	1.4944
Heavy barium crown	1.6130
Lubricating oil (automobile)	1.50
Water	1.33

TABLE 20.—DENSITIES

1 kg./m.^{3} = 10^{-3} g./cm.^{3} = 1 g./l.

Substance	kg./m.^{3}	°C.
Air	1.205	20 and 0.76 m. of Hg pressure
Alcohol	789	20
Aluminum	2,700	20
Amber	1,000–1,100	20
Cast iron	7,250	20
Copper	8,940	20
Cork	240	20
Diamond	3,510	20
Gasoline	660– 690	20
Glass	2,400–2,600	20
Glycerin	1,260	20
Graphite	2,260	20
Hydrogen	0.08987	0 and 0.76 m. of Hg pressure
Ice	880– 920	20
Iron	7,860	20
Steel	7,800	20
Ivory	1,830–1,920	20
Lead	11,350	20
Leather	1,020	20
Mercury	13,546	20
Nickel	8,900	20
Paraffin	870– 910	20
Platinum	21,450	20
Rock salt	2,163	20
Rubber	910– 930	20
Sea water	1,025	20
Silver	10,500	20
Steam	0.589	100 and 0.76 m. of Hg pressure
Water	998.203	20
	958	100
Wood	310–1,330	

TABLE 21.—NUMERICAL CONSTANTS AND FORMULAS

$\pi = 3.1416$ $\qquad \pi^2 = 9.8696$ $\qquad 1/\pi = 0.31831$

$e = 2.7183$ = the base of the natural or Napierian logarithms

$\log_e x = (\log_{10} x)(\log_e 10) = 2.30259 \log_{10} x$

Area of a circle of radius $r = \pi r^2$

Area of a sphere of radius $r = 4\pi r^2$

Volume of a sphere of radius $r = 4\pi r^3/3$

TABLE 22.—GENERAL PHYSICAL AND ASTROPHYSICAL CONSTANTS

Avogadro's number	$N = 6.03 \times 10^{23}$ molecules/mole
Charge on the electron	$e = 16.0 \times 10^{-20}$ coulomb
Faraday	= 96,489 coulombs/chemical equivalent
Mass of the electron	$m = 9.09 \times 10^{-31}$ kg.
Mass of the hydrogen atom	$m_H = 1.66 \times 10^{-27}$ kg.
Mechanical equivalent of heat	$J = 4.185$ joules/cal.
Molar gas constant	$R = 8.3136$ joules/mole °K.
Boltzmann's constant	$k = 1.38 \times 10^{-23}$ joule/molecule °K.
Planck constant	$h = 6.55^* \times 10^{-34}$ joule-sec.
Ratio of charge to mass of electron	$e/m = 1.76 \times 10^{11}$ coulombs/kg.
Ratio of charge to mass of proton	$e/m_H = 9.564 \times 10^{7}$ coulombs/kg.
Velocity of light	2.99796×10^{8} m./sec.
Velocity of sound	1,129 ft./sec. at 20°C.
Permeability for free space	$\mu_0 = 4\pi \times 10^{-7}$ newton/amp.^{2}
Dielectric constant for a vacuum	$\epsilon_0 = 1/(4\pi \times 8.98776 \times 10^{9})$ farads/m.
	$= 8.84 \times 10^{-12}$
Acceleration of gravity	earth, $g = 9.8$ m./sec.^{2}
	Jupiter 24.3 m./sec.^{2}
	moon 1.55 m./sec.^{2}
	sun 273 m./sec.^{2}
Gravitational constant	$G = 6.66 \times 10^{-11}$ newton-m.^{2}/kg.^{2}
	$= 66.66 \times 10^{-8}$ dyne cm.^{2}/g.^{2}
	$= 1.0669 \times 10^{-9}$ poundal ft.^{2}/lb.^{2}
Mass	earth 5.95×10^{24} kg.
	Jupiter 1.89×10^{27} kg.
	moon 7.3×10^{22} kg.
	sun 1.985×10^{30} kg.
Radius	earth 6.367×10^{6} m.
	Jupiter 7.13×10^{7} m.
	moon 1.74×10^{6} m.
	sun 6.96×10^{8} m.
Radius of orbit	earth 1.495×10^{11} m.
	moon 3.85×10^{8} m.
	Jupiter 7.76×10^{11} m.

*6.61×10^{-34} is probably more accurate.

II. CONVERSION TABLES

TABLE 23.—LENGTH CONVERSION FACTORS

	Kilometers	Meters	Centimeters	Miles	Feet	Inches
1 km.	1	10^{3}	10^{5}	0.62137	3.281×10^{3}	3.937×10^{4}
1 m.	10^{-3}	1	100	6.2137×10^{-4}	3.281	39.37
1 cm.	10^{-5}	10^{-1}	1	6.2137×10^{-6}	3.281×10^{-2}	0.3937
1 mi.	1.6093	1609	1.609×10^{5}	1	5280	6.336×10^{4}
1 ft.	3.048×10^{-4}	0.3048	30.48	1.894×10^{-4}	1	12
1 in.	2.54×10^{-5}	2.54×10^{-2}	2.54	1.5783×10^{-4}	0.8333	1

TABLE 24.—MASS CONVERSION FACTORS

1 pound = 0.45359 kilogram
1 kilogram = 2.20462 pounds

TABLE 25.—FORCE CONVERSION FACTORS

	Newton	Dyne	Kg. wt	G. wt.	Poundal	Lb. wt.	Ton wt.
1 newton	1	10^5	0.10198	101.98	7.2335	0.224827	1.1241×10^{-4}
1 dyne	10^{-5}	1	1.0198×10^{-1}	1.0198×10^{-3}	7.2335×10^{-5}	2.24827×10^{-6}	1.1241×10^{-9}
1 kg. wt.	9.806	9.806×10^5	1	10^3	70.931	2.20462	1.10231×10^{-3}
1 g. wt.	9.806×10^{-3}	980.6	10^{-3}	1	0.070931	2.20462×10^{-3}	1.10231×10^{-6}
1 poundal	0.13825	1.3825×10^4	0.014098	14.098	1	0.031081	1.55405×10^{-5}
1 lb. wt.	4.448	4.448×10^5	0.45359	453.59	32.174	1	5×10^{-4}
1 ton wt.	8896	8.896×10^{-8}	907.18	9.0718×10^5	6.4348×10^4	2000	1

TABLE 26.—PRESSURE CONVERSION FACTORS

	Atm.	Newton/m.²	Dyne/cm.²	Lb./in.²	Cm. Hg
1 atm.	1	1.0132×10^5	1.0132×10^6	14.696	76.0
1 newton/m.²	9.8697×10^{-6}	1	10	1.4504×10^{-4}	7.5010×10^{-4}
1 dyne/cm.²	9.8697×10^{-7}	10^{-1}	1	1.4504×10^{-5}	7.5010×10^{-5}
1 lb./in.²	6.8046×10^{-2}	6,894.4	68,944	1	5.1715
1 cm. Hg	1.3158×10^{-2}	1,332	13,320	1.9337×10^{-1}	1

TABLE 27.—ENERGY CONVERSION FACTORS

	Joule	Kg.-m.	Erg	Ft.-lb.	Ft.-poundal	Cal.	B. t. u.	Electron volts	Mass units	L.-atm.
1 joule	1	0.1019	10^7	0.73756	23.73	0.2389	9.482×10^{-4}	6.255×10^{18}	3.715×10^{9}	9.689×10^{-3}
1 kg.-m	9.806	1	9.806×10^7	7.2327	232.7	2.343	9.298×10^{-3}	6.134×10^{19}	6.585×10^{10}	9.678×10^{-2}
1 erg	10^{-7}	1.02×10^{-8}	1	7.38×10^{-8}	2.37×10^{-6}	2.39×10^{-8}	9.482×10^{-11}	6.255×10^{11}	6.715×10^{2}	9.689×10^{-10}
1 ft.-lb	1.356	0.13827	1.356×10^7	1	32.174	0.3240	1.286×10^{-3}	8.484×10^{18}	9.107×10^{9}	1.338×10^{-2}
1 ft.-poundal	0 04215	4.297×10^{-3}	4.215×10^5	0.031081	1	0.01007	3.997×10^{-5}	2.637×10^{17}	$2\ 831 \times 10^{8}$	4.084×10^{-4}
1 cal	4.185	0.427	4.185×10^7	3.09	99.3	1	3.968×10^{-3}	2.618×10^{19}	2.810×10^{10}	4.130×10^{-2}
1 B. t. u	1,055	107.5	1.055×10^{10}	778	2.5×10^4	252	1	6.597×10^{21}	7.082×10^{12}	10.41
1 electron volt	1.60×10^{-19}	1.629×10^{-20}	1.60×10^{-12}	1.179×10^{-19}	3.794×10^{-18}	3.819×10^{-20}	1.5159×10^{-22}	1	1.07356×10^{-9}	1.55×10^{-21}
1 mass unit	1.49×10^{-10}	1.517×10^{-11}	1.49×10^{-3}	1.098×10^{-10}	$3\ 534 \times 10^{-9}$	3.557×10^{-11}	1.4120×10^{-13}	9.31480×10^{8}	1	1.44×10^{-12}
1 l.-atm.	101.32	10.333	1.013×10^9	74.733	2.404×10^3	24.21	9.607×10^{-2}	6.338×10^{20}	6.804×10^{11}	1

TABLE 28.—RELATIONS BETWEEN THREE SYSTEMS OF ELECTRICAL UNITS
Equality signs are implied across any row, *i.e.*, 1 m. = 100 cm.

Entity	Absolute practical	Electrostatic	Electromagnetic
Length	1 m.	100 cm.	100 cm.
Mass	1 kg.	1,000 g.	1,000 g.
Time	1 sec.	1 sec.	1 sec.
Force	1 newton	10^5 dynes	10^5 dynes
Work, Energy	1 joule	10^7 ergs	10^7 ergs
Power	1 watt	10^7 ergs/sec.	10^7 ergs/sec.
Charge	1 coulomb	3×10^9	10^{-1}
Current	1 amp.	3×10^9	10^{-1}
Electric field	1 volt/m.	$\frac{1}{(3 \times 10^4)}$	10^6
Electromotive force or potential difference	1 volt	$1/300$	10^8
Resistance	1 ohm	$\frac{1}{(9 \times 10^{11})}$	10^9
Capacity	1 farad	9×10^{11}	10^{-9}
Flux	1 weber	$1/300$	10^8 maxwells
Magnetic induction	1 weber/m.^{2}	$\frac{1}{(3 \times 10^6)}$	10^4 gausses
Magnetic field	1 amp. turn/m.	$12\pi \times 10^7$	$4\pi \times 10^{-3}$ oersted
Inductance	1 henry	$\frac{1}{(9 \times 10^{11})}$	10^9
Pole strength	1 weber	$\frac{1}{(12\pi \times 10^2)}$	$\frac{10^8}{4\pi}$
Permeability of free space	$4\pi \times 10^{-7}$ henry/m.	$\frac{1}{(9 \times 10^{20})}$(sec./cm.)^2	unity
Dielectric constant of free space	$\frac{1}{4\pi \times 9 \times 10^9}$farad/m.	unity	$\frac{1}{(9 \times 10^{20})}$sec.^{2}/cm.^{2}

In the preceding table c, the velocity of light, has been taken as 3×10^8 m./sec.

INDEX

A

N